TITLE

METHOD FOR THE PRODUCTION OF *p*-HYDROXYBENZOATE IN SPECIES OF *PSEUDOMONAS* AND *AGROBACTERIUM*

This application is a continuation in part of U.S. Application 09/585,174 filed on June 1, 2000.

FIELD OF INVENTION

The present invention relates to the fields of molecular biology and microbiology, and to the use of genetic techniques to introduce a modified pathway for the production of desired compounds. More specifically, this invention describes genetically engineered biocatalysts possessing an enhanced, or new, ability to transform *p*-cresol or toluene to *p*-hydroxybenzoate.

BACKGROUND OF THE INVENTION

p-Hydroxybenzoate (PHBA) is used as a monomer for synthesizing Liquid Crystal Polymers (LCP). LCP's are used in electronic connectors and in telecommunication and aerospace applications. LCP resistance to sterilizing radiation suits these materials for use in medical devices as well as in chemical, and food packaging applications. Esters of PHBA also are used as backbone modifiers in other condensation polymers (i.e., polyesters), and are also used to make parabens preservatives.

Chemical synthesis of PHBA is known. For example, JP 05009154 teaches a chemical route using the Kolbe-Schmidt process from tar acid and CO_2 involving 1) the extraction of tar acid from a tar naphthalene oil by an aqueous potassium hydroxide, 2) adding phenol to the extracted tar acid potassium salt, 3) removing H_2O, and 4) reacting the resultant slurry with CO_2. Alternative methods of chemical synthesis are known (see, for example, U.S. 5,399,178; U.S. 4,740,614; and U.S. 3,985,797).

However, chemical synthesis of PHBA is problematic and costly due to the high energy needed for synthesis and the extensive purification of product required. An alternate low cost method with simplified purification would represent an advance in the art. Biological production offers one such low cost, simplified solution to this problem.

Microbiological methods of PHBA synthesis are known. For example, JP 06078780 teaches PHBA preparation by culturing benzoic acid in the presence of microorganisms (preferably *Aspergillus*) that oxidize benzoic acid to PHBA.

An alternate method of biological production is suggested by bacteria that have an enzymatic pathway for the degradation of toluene

and other organics where PHBA is produced as an intermediate. The first enzyme in the toluene degradation pathway is toluene monooxygenase (TMO) and the pathway is referred to as the TMO pathway. The steps of the TMO pathway have been described (Whited and Gibson, *J. Bacteriol.* 173:3010-3020 (1991)) and are illustrated in Figure 1. Bacteria that possess the toluene degradation pathway are found in the genus *Pseudomonas* where *Pseudomonas putida*, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa* and *Pseudomonas mendocina* are the most commonly utilized species. Other examples of aerobic bacteria that are known to degrade toluene are *Burkholderia* (Johnson et al., *Appl. Environ. Microbiol.* 63:4047-4052 (1997)), *Mycobacterium* (Stephen et al., *Appl. Environ. Microbiol.* 64:1715-1720 (1998)), *Sphingomonas* (Zylstra et al., *J. Ind. Microbiol. Biotechnol.* 19:408-414 (1997)) and *Rhodococcus* (Kosono et al., *Appl. Environ. Microbiol.* 63:3282-3285 (1997)). In addition, several different species of anaerobic bacteria are known to utilize toluene (Heider et al., *Anarobe* 3:1-22 (1997)). Toluene degradation pathways have been highly characterized (Romine et al., In *Bioremediation of Chlorinated Polycyclic Aromatic Hydrocarbon Compounds*; Hinchee, R. E., Ed.; Lewis: Boca Raton, FL, 1994; pp 271-276) and a number of the genes encoding key enzymes have been cloned and sequenced, including the protocatechuate 3,4-dioxygenase genes (Frazee, *J. Bacteriol.* 175(19):6194-6202 (1993)), the *pcaR* regulatory gene from *Pseudomonas putida*, which is required for the complete degradation of *p*-hydroxybenzoate (Romero-Steiner et al., *J. Bacteriol.* 176(18):5771-5779 (1994); Dimarco et al., *J. Bacteriol.* 176(14):4277-4284 (1994)) and the *pobA* gene encoding the expression of *p*-hydroxybenzoate hydroxylase (PHBH), the principal enzyme for the conversion of PHBA to protocatechuate (Wong et al., *Microbiology* (Reading U.K.) 140(10):2775-2786 (1994); Entsch et al., *Gene* 71(2):279-291 (1988)).

Bacteria that possess the TMO pathway are useful for degrading toluene and trichloroethylene. They are able to use these and other organics as sole carbon sources where they are transformed through PHBA to ring-opening degradation products (U.S. 5,017,495; U.S. 5,079,166; U.S. 4,910,143). By using the chromosomal TMO pathway, in combination with mutations that prevent PHBA degradation in *Pseudomonas mendocina* KR1, it has been shown that PHBA can be accumulated by oxidation of toluene (PCT/US98/12072).

Recently, various strains of *Pseudomonas* possessing the toluene degradation pathways have been used to produce muconic acid *via* manipulation of growth conditions (U.S. 4,657,863; U.S. 4,968,612). Additionally, strains of *Enterobacter* with the ability to convert *p*-cresol to PHBA have been isolated from soil (JP 05328981). Further, JP 05336980 and JP 05336979 disclose isolated strains of *Pseudomonas putida* with the ability to produce PHBA from *p*-cresol. Additionally, Miller and coworkers (*Green Chem.* 1(3):143-152 (1999)) have shown the bioconversion of toluene to PHBA via the construction of a recombinant *Pseudomonas putida*. Their initial catalyst development focused on *Pseudomonas mendocina* KR1 for production of PHBA from toluene. However, they were unable to obtain significant accumulation of PHBA from toluene using this strain. This result was due to their inability to obtain a sufficient disruption of PobA activity (the enzyme catalyzing *m*-hydroxylation of PHBA to protocatechuate in the protocatechuate branch of the β-ketoadipate pathway; see Figure 1).

Although the presence of the TMO pathway in *Pseudomonas mendocina* KR1 has been documented (Wright and Olsen, *Applied Environ. Microbiol.* 60(1):235-242 (1994)), the art has not provided a molecular characterization and sequence of the *pcu* genes encoding the enzymes that transform *p*-cresol to PHBA in this organism. The art has also not provided bacterial host cells harboring novel recombinant plasmids encoding the enzymes of *p*-cresol to PHBA oxidation, together with operably-linked native promoter and regulatory sequences and proteins. Such bacterial host strains, if they lack the enzymes to degrade PHBA further, can accumulate PHBA when cultured in the presence of *p*-cresol.

As an alternative to culturing cells in the presence of *p*-cresol, the latter compound can be formed from toluene in cells that additionally harbor plasmid-encoded toluene monooxygenase. A bacterial strain harboring plasmid-encoded *tmo* and *pcu* operons has not been fully described in the art, particularly a strain that exceeds the production level of PHBA when compared to plasmid-free *Pseudomonas mendocina* KR1. In addition, expression of the *tmo* operon using its native toluene-induced promoter localized upstream of a *tmoX* gene previously has not been known. Therefore, the problem to be solved is the lack of a fully characterized *pcu* operon and the availability of a bacterial strain harboring

plasmid-encoded *tmo* and *pcu* operons to use for the bioproduction of PHBA.

SUMMARY OF THE INVENTION

The present invention solves the problem of extensively characterizing the *pcu* operon by providing cloned, sequenced, and expressed genes of the *pcu* operon from *Pseudomonas mendocina* KR-1 that can be transformed into and used to produce PHBA from *p*-cresol in *Pseudomonas putida* and *Agrobacterium rhizogenes* strains that do not normally possess this capability. In addition, transformation of the *pcu* operon into *Pseudomonas mendocina* KRC16KDpobA51 supplements the endogenous *pcu* operon leading to an increase in PHBA production. This increase in PHBA production in *Pseudomonas mendocina* KRC16KDpobA51 transformed with plasmid-encoded *pcu* is an improvement over PCT/US98/12072.

The present invention provides a method for the production of PHBA comprising: (i) culturing a *Pseudomonas*, *Agrobacterium* or related strain transformed with a *pcu* operon in a medium containing an aromatic organic substrate, at least one suitable fermentable carbon source, and a nitrogen source, wherein the supplied *pcu* operon comprises genes encoding the TMO toluene degradation pathway enzymes *p*-cresol methylhydroxylase and *p*-hydroxybenzaldehyde dehydrogenase, the transcriptional activator PcuR, wherein the transformed *Pseudomonas* or *Agrobacterium* strain does not produce any detectable *p*-hydroxybenzoate hydroxylase, whereby PHBA accumulates; and (ii) recovering the PHBA.

The present invention also encompasses the combination of the *pcu* and *tmo* operons on a single replicon such that expression of *tmo* is obtained by transcription from a previously undisclosed toluene or *p*-cresol induced *tmoX* promoter, and expression of *pcu* is obtained by transcription using a previously undisclosed sequence encoding a transcriptional activator.

Another preferred embodiment of the present invention includes the recombinant plasmid pMC4 in *Pseudomonas putida* DOT-T1. This strain synthesized the highest levels of *tmo* and *pcu*-encoded enzymes observed and is described herein.

It has also been found that the heterologous todST proteins that control the induction of toluene dioxygenase pathway induce high levels of expression from the *tmo* pathway genes, and are useful tools to mediate

expression of the catabolic *tmo* genes and PHBA production in any organism that does not possess these genes.

BRIEF DESCRIPTION OF THE DRAWINGS, SEQUENCE DESCRIPTIONS, AND BIOLOGICAL DEPOSITS

The invention can be more fully understood with reference to the drawings, from the detailed description, and the sequence descriptions which form part of this application.

Figure 1 illustrates the pathway of the toluene degradation in *Pseudomonas mendocina* KR-1.

Figure 2 illustrates *Pseudomonas mendocina* KR-1 *pcu* operon.

Figure 3 illustrates *Pseudomonas mendocina* KR-1 *tmo* operon.

Figure 4 illustrates the *pcu* and *tmo* expression plasmid pMC4.

Figure 5 illustrates pMIR30 and the strategy used for its construction.

Figure 6 illustrates pMIR31 and the strategy used for its construction.

Figure 7 illustrates pMIR44 and the strategy used for its construction.

Figure 8 illustrates the cluster of *tmo/pcu/pobA* genes in *Pseudomonas mendocina* KR1.

The following 142 sequence descriptions contained in the sequences listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825 ("Requirements for Patent Applications Containing Nucleotide Sequences and/or Amino Acid Sequence Disclosures – the Sequence Rules") and are consistent with World Intellectual Property Organization (WIPO) Standard ST2.5 (1998) and the sequence listing requirements of the EPO and PCT (Rules 5.2 and 49.5(a-bis), and Section 208 and Annex C of the Administration Instructions). The Sequence Descriptions contain the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IYUB standards described in *Nucleic Acids Res.* 13:3021-3030 (1985) and in the *Biochemical Journal* 219:345-373 (1984) which are herein incorporated by reference.

SEQ ID NO:1 is the nucleotide sequence of the *pcu* operon isolated from *Pseudomonas mendocina* KR-1 (6491 bp).

SEQ ID NO:2 is the deduced amino acid sequence of the transcriptional activator PcuR encoded by ORF1.1 (SEQ ID NO:98).

SEQ ID NO:3 is the deduced amino acid sequence of PcuC encoded by ORF1.2 (SEQ ID NO:99) which has the enzyme activity of PHBAD.

SEQ ID NO:4 is the deduced amino acid sequence of PcuA encoded by ORF1.3 (SEQ ID NO:100) which has the enzyme activity of PCMH.

SEQ ID NO:5 is the deduced amino acid sequence of PcuX encoded by ORF1.4 (SEQ ID NO:101) which is an unidentified open reading frame and which may be an inner membrane protein.

SEQ ID NO:6 is the predicted amino acid sequence of PcuB encoded by ORF1.5 (SEQ ID NO:102) which has the enzyme activity of PCMH.

SEQ ID NOs:7-77 are the nucleotide sequences of primers used for sequencing *pcu*.

SEQ ID NOs:78-79 are the nucleotide sequences of primers used for cloning a *Pseudomonas putida* (NCIMB 9869) *pchC* gene.

SEQ ID NOs:80-90 are the nucleotide sequences of primers used for sequencing *tmoX*.

SEQ ID NO:91 is the nucleotide sequence of the *tmoX* gene and its 5' non-translated region from *Pseudomonas mendocina* KR-1.

SEQ ID NO:92 is the deduced amino acid sequence of TmoX encoded by ORF2.1 (SEQ ID NO:103).

SEQ ID NOs:93-94 are the nucleotide sequences of primers used for cloning *pcu* for insertion into pMC3.

SEQ ID NOs:95-96 are the nucleotide sequence of primers used for constructing plasmids pPCUR1 and pPCUR2.

SEQ ID NO:97 is the nucleotide sequence of the primer used to map the transcript initiation site of *tmoX*.

SEQ ID NO:98 is the nucleotide sequence of the transcriptional activator PcuR (ORF1.1).

SEQ ID NO:99 is the nucleotide sequence of PcuC (ORF1.2).

SEQ ID NO:100 is the nucleotide sequence of PcuA (ORF1.3).

SEQ ID NO:101 is the nucleotide sequence of PcuX (ORF1.4).

SEQ ID NO:102 is the nucleotide sequence of PcuB (ORF1.5).

SEQ ID NO:103 is the nucleotide sequence of the *tmoX* gene from *Pseudomonas mendocina* KR-1 (ORF2.1).

SEQ ID NO:104 is a primer used to identify the *pobA* gene.

SEQ ID NO:105 is a primer used to identify the *pobA* gene.

SEQ ID NO:106 is a primer used to identify the *pobA* gene.

SEQ ID NO:107 is a primer used to identify the *pobA* gene.

SEQ ID NO:108 is a primer used to identify the *pobA* gene.

SEQ ID NO:109 is a primer used to identify the *pobA* gene.

SEQ ID NO:110 is a primer used to identify the *pobA* gene.

SEQ ID NO:111 is a primer used to identify the *pobB* gene.

SEQ ID NO:112 is the nucleotide sequence of the *todST* genes.

SEQ ID NO:113 is the nucleotide sequence of the *tmoST* genes (4821 bp).

SEQ ID NO:114 is the nucleotide sequence of the *tmoS* gene.

SEQ ID NO:115 is the nucleotide sequence of the *tmoT* gene.

SEQ ID NO:116 is the deduced amino acid sequence of TmoS encoded by SEQ ID NO:114 which has the enzyme activity of bZIP histidine kinase.

SEQ ID NO:117 is the deduced amino acid sequence of TmoT encoded by SEQ ID NO:115 which has the activity of a response regulator.

SEQ ID NO:118-142 are the primer sequences used to sequence *tmoST* genes in pMIR60 (pUC18*Hind*III::insert *Hind*III of pMAX47-2)

Applicant has made the following biological deposits under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for the Purposes of Patent Procedure:

Depositor Identification Reference	International Depository Designation	Date of Deposit
Pseudomonas mendocina KRC16KDpobA51	ATCC 55885	11/26/96

The Depositor has authorized the Applicant to refer to the deposited material in the application and has given his unreserved and irrevocable consent to the deposited material being made available to the public in accordance with Rule 28 of the Implementing Regulations to the European Patent Convention (Rule 28(1)(d) EPC).

DETAILED DESCRIPTION OF THE INVENTION

PHBA is a valuable monomer for the synthesis of liquid crystalline polymers (LCP). Applicants have provided methods for the biological production of PHBA from genetically engineered *Pseudomonas*, *Agrobacterium*, or related strains transformed with a pcu operon. The instant methods provide PHBA without the high energy cost of synthetic production and without producing toxic waste streams. Applicants have also provided a method for the biological production of *p*-cresol from genetically engineered *Escherichia* or *Pseudomonas*.

The following abbreviations and definitions will be used to interpret the specification and the claims.

"*para*-Hydroxybenzoic acid", "*para*-hydroxybenzoate", "*p*-hydroxybenzoate" or "4-hydroxybenzoic acid" is abbreviated PHBA.

"*para*-Hydroxybenzoate hydroxylase" is abbreviated PHBH.

"Toluene-4-monooxygenase" is abbreviated TMO.

"*para*-Cresol methylhydroxylase" is abbreviated PCMH.

"*para*-Hydroxybenzaldehyde dehydrogenase" is abbreviated PHBAD.

" protocatechuate 4,5-dioxygenase β-chain" is abbreviated ligB.

"TnpA2* and TnpA3*" are abbreviated tnp.

"Ethylenediaminetetraacetic acid" is abbreviated EDTA.

"Isopropyl-β-D-thiogalactopyranoside" is abbreviated IPTG.

"Shrimp alkaline phosphatase" is abbreviated SAP.

"Calf intestinal alkaline phosphatase" is abbreviated CIP.

"Phenazine ethosulfate" is abbreviated PES.

"2,6-Dichlorophenol-indophenol" is abbreviated DCPIP.

"SSC" is the abbreviation for 150 mM NaCl, 15 mM sodium citrate, pH 7.0.

"TE" is the abbreviation for 10 mM Tris-HCl, 1 mM EDTA, pH 8.0.

The term "amp" refers to ampicillin.

The term "chl" refers to chloramphenicol.

The term "kan" refers to kanamycin.

The term "strep" refers to streptomycin.

The term "Pip" refers to peperacillan.

The term "tet" refers to tetracycline.

The term "strR" refers to a gene conferring resistance to streptomycin.

The terms "TMO degradative pathway" or "TMO enzymatic pathway" refer to the enzymes and genes encoding the enzymes found in some *Pseudomonas* bacteria that are responsible for the degradation of toluene, *p*-cresol and similar aromatic substrates. The TMO pathway is outlined in Figure 1 and contains at least toluene-4-monooxygenase (TMO), *p*-cresol methylhydroxylase (PCMH), *p*-hydroxybenzoaldehyde dehydrogenase (PHBAD), and *p*-hydroxybenzoate hydroxylase (PHBH).

The term "aromatic organic substrate" refers to an aromatic compound that is degraded by the TMO enzymatic pathway. Typical examples of suitable aromatic substrates are toluene, *p*-cresol, p-hydroxybenzyl, and p-hydroxybenzaldehyde.

TmoST gene refers to genes identified as tmoS and tmoT. These genes encode for TmoS and TmoT (TmoST) polypeptides (proteins). TmoS and TmoT proteins fucntion as transcriptional activators of the promoter *tmoXABCDEF* operon.

The terms "plasmid", "vector" and "cassette" refer to an extra chromosomal element often carrying genes which are not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA molecules. Such elements may be autonomously replicating sequences, genome-integrating sequences, phage or nucleotide sequences, linear or circular, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell. "Transformation vector" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that facilitate transformation of a particular host cell.

An "isolated nucleic acid molecule" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid molecule in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

"Substantially similar" refers to nucleic acid molecules wherein changes in one or more nucleotide bases result in substitution of one or more amino acids, but do not affect the functional properties of the protein encoded by the DNA sequence. "Substantially similar" also refers to nucleic acid molecules wherein changes in one or more nucleotide bases

do not affect the ability of the nucleic acid molecule to mediate alteration of gene expression by antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid molecules of the instant invention (such as deletion or insertion of one or more nucleotide bases) that do not substantially affect the functional properties of the resulting transcript vis-à-vis the ability to mediate alteration of gene expression by antisense or co-suppression technology or alteration of the functional properties of the resulting protein molecule. The invention encompasses more than the specific exemplary sequences.

For example, it is well known in the art that alterations in a gene which result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded protein are common. For the purposes of the present invention substitutions are defined as exchanges within one of the following five groups:

1. Small aliphatic, nonpolar or slightly polar residues: Ala, Ser, Thr (Pro, Gly);
2. Polar, negatively charged residues and their amides: Asp, Asn, Glu, Gln;
3. Polar, positively charged residues: His, Arg, Lys;
4. Large aliphatic, nonpolar residues: Met, Leu, Ile, Val (Cys); and
5. Large aromatic residues: Phe, Tyr, Trp.

Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue (such as glycine) or a more hydrophobic residue (such as valine, leucine, or isoleucine). Similarly, changes which result in substitution of one negatively charged residue for another (such as aspartic acid for glutamic acid) or one positively charged residue for another (such as lysine for arginine) can also be expected to produce a functionally equivalent product.

In many cases, nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the protein molecule would also not be expected to alter the activity of the protein.

Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Moreover, the skilled artisan recognizes that substantially similar sequences encompassed by this invention are also defined by their

ability to hybridize, under stringent conditions (0.1X SSC, 0.1% SDS, 65°C and washed with 2X SSC, 0.1% SDS followed by 0.1X SSC, 0.1% SDS), with the sequences exemplified herein. Preferred substantially similar nucleic acid fragments of the instant invention are those nucleic acid fragments whose DNA sequences are at least 80% identical to the DNA sequence of the nucleic acid fragments reported herein. More preferred nucleic acid fragments are at least 90% identical to the DNA sequence of the nucleic acid fragments reported herein. Most preferred are nucleic acid fragments that are at least 95% identical to the DNA sequence of the nucleic acid fragments reported herein.

A nucleic acid fragment is "hybridizable" to another nucleic acid fragment, such as a cDNA, genomic DNA, or RNA, when a single stranded form of the nucleic acid fragment can anneal to the other nucleic acid fragment under the appropriate conditions of temperature and solution ionic strength. Hybridization and washing conditions are well known and exemplified in Sambrook, J., Fritsch, E. F. and Maniatis, T. Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989), particularly Chapter 11 and Table 11.1 therein (entirely incorporated herein by reference). The conditions of temperature and ionic strength determine the "stringency" of the hybridization. For preliminary screening for homologous nucleic acids, low stringency hybridization conditions, corresponding to a T*m* of 55 °, can be used, e.g., 5X SSC, 0.1% SDS, 0.25% milk, and no formamide; or 30% formamide, 5X SSC, 0.5% SDS. Moderate stringency hybridization conditions correspond to a higher T*m*, e.g., 40% formamide, with 5X or 6X SSC. Hybridization requires that the two nucleic acids contain complementary sequences, although depending on the stringency of the hybridization, mismatches between bases are possible. The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and the degree of complementation, variables well known in the art. The greater the degree of similarity or homology between two nucleotide sequences, the greater the value of T*m* for hybrids of nucleic acids having those sequences. The relative stability (corresponding to higher T*m*) of nucleic acid hybridization decreases in the following order: RNA:RNA, DNA:RNA, DNA:DNA. For hybrids of greater than 100 nucleotides in length, equations for calculating T*m* have been derived (see Sambrook et al., *supra*, 9.50-9.51). For hybridization with shorter nucleic acids, i.e., oligonucleotides, the position of mismatches becomes more

important, and the length of the oligonucleotide determines its specificity (see Sambrook et al., *supra*, 11.7-11.8). In one embodiment the length for a hybridizable nucleic acid is at least about 10 nucleotides. Preferable a minimum length for a hybridizable nucleic acid is at least about 15 nucleotides; more preferably at least about 20 nucleotides; and most preferably the length is at least 30 nucleotides. Furthermore, the skilled artisan will recognize that the temperature and wash solution salt concentration may be adjusted as necessary according to factors such as length of the probe.

A "substantial portion" refers to an amino acid or nucleotide sequence which comprises enough of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to afford putative identification of that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul et al., *J. Mol.Biol.* 215:403-410 (1993); see also www.ncbi.nlm.nih.gov/BLAST/). In general, a sequence of ten or more contiguous amino acids or thirty or more nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene specific oligonucleotide probes comprising 20-30 contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., in situ hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as amplification primers in PCR in order to obtain a particular nucleic acid molecule comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises enough of the sequence to afford specific identification and/or isolation of a nucleic acid molecule comprising the sequence. The instant specification teaches partial or complete amino acid and nucleotide sequences encoding one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for the purpose known to those skilled in the art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

The term "complementary" describes the relationship between nucleotide bases that are capable to hybridizing to one another. For example, with respect to DNA, adenosine is complementary to thymine and cytosine is complementary to guanine. Accordingly, the instant invention also includes isolated nucleic acid molecules that are complementary to the complete sequences as reported in the accompanying Sequence Listing as well as those substantially similar nucleic acid sequences.

The term "percent identity", as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in: *Computational Molecular Biology*; Lesk, A. M., Ed.; Oxford University Press: New York, 1988; *Biocomputing: Informatics and Genome Projects*; Smith, D. W., Ed.; Academic Press: New York, 1993; *Computer Analysis of Sequence Data, Part I*; Griffin, A. M. and Griffin, H. G., Eds.; Humana Press: New Jersey, 1994; *Sequence Analysis in Molecular Biology*; von Heinje, G., Ed.; Academic Press: New York, 1987; and *Sequence Analysis Primer*; Gribskov, M. and Devereux, J., Eds.; Stockton Press: New York, 1991. Preferred methods to determine identity are designed to give the largest match between the sequences tested.

Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG Pileup program found in the GCG program package, using the Needleman and Wunsch algorithm with their standard default values of gap creation penalty=12 and gap extension penalty=4 (Devereux et al., *Nucleic Acids Res.* 12:387-395 (1984)), BLASTP, BLASTN, and FASTA (Pearson et al., *Proc. Natl. Acad. Sci. USA* 85:2444-2448 (1988). The BLASTX program is publicly available from NCBI and other sources (BLAST Manual, Altschul et al., Natl. Cent. Biotechnol. Inf., Natl. Library Med. (NCBI NLM) NIH, Bethesda, Md. 20894; Altschul et al., *J. Mol. Biol.* 215:403-410 (1990); Altschul et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database

search programs", *Nucleic Acids Res.* 25:3389-3402 (1997)). Another preferred method to determine percent identity, is by the method of DNASTAR protein alignment protocol using the Jotun-Hein algorithm (Hein et al., *Methods Enzymol.* 183:626-645 (1990)). Default parameters for the Jotun-Hein method for alignments are: for multiple alignments, gap penalty=11, gap length penalty=3; for pairwise alignments ktuple=6. As an illustration, by a polynucleotide having a nucleotide sequence having at least, for example, 95% "identity" to a reference nucleotide sequence it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. Analogously, by a polypeptide having an amino acid sequence having at least, for example, 95% identity to a reference amino acid sequence is intended that the amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

The term "percent homology" refers to the extent of amino acid sequence identity between polypeptides. When a first amino acid sequence is identical to a second amino acid sequence, then the first and second amino acid sequences exhibit 100% homology. The homology between any two polypeptides is a direct function of the total number of matching amino acids at a given position in either sequence, e.g., if half of the total number of amino acids in either of the two sequences are the same then the two sequences are said to exhibit 50% homology.

"Codon degeneracy" refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid molecule that encodes all or a substantial portion of the amino acid sequence encoding the PcuR, PcuC, PcuA, PcuX and PcuB proteins as set forth in SEQ ID NO:2 through SEQ ID NO:6, and also to any nucleic acid molecule that encodes all or a substantial portion of the amino acid sequence encoding the TmoX protein as set forth in SEQ ID NO:92. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a gene for improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

"Synthetic genes" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form gene segments which are then enzymatically assembled to construct the entire gene. "Chemically synthesized", as related to a sequence of DNA, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of DNA may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the genes can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

"Gene" refers to a nucleic acid molecule that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers to any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

"Coding sequence" refers to a DNA sequence that codes for a specific amino acid sequence.

"Suitable regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

"Promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. An "enhancer" is a DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues

or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA molecules of different lengths may have identical promoter activity.

The "translation leader sequence" refers to a DNA sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner et al., *Mol. Biotech.* 3:225 (1995)).

The "3' non-coding sequences" refer to DNA sequences located downstream of a coding sequence and include recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al., *Plant Cell* 1:671-680 (1989).

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a double-stranded DNA that is complementary to and derived from mRNA. "Sense" RNA refers to RNA transcript that includes the mRNA and so can be translated into protein by the cell. "Antisense RNA" refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (U.S. Patent No. 5,107,065). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to antisense RNA, ribozyme

RNA, or other RNA that is not translated yet and has an effect on cellular processes.

The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid molecule so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it affects the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid molecule of the invention. Expression may also refer to translation of mRNA into a polypeptide. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020).

"Altered levels" refers to the production of gene product(s) in organisms in amounts or proportions that are not characteristic of normal, wild-type, or non-transformed organisms. The altered level may be either an increase or decrease in the amount or proporiton of gene product relative to that produced by the normal, wild-type, or non-transformed organism.

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed. "Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

A "fragment" constitutes a fraction of the DNA sequence of the particular region.

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid

fragments are referred to as "transgenic" or "recombinant" or "transformed" organisms.

The terms "plasmid", "vector" and "cassette" refer to an extra chromosomal element often carrying genes which are not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA fragments. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear or circular, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell. "Transformation cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that facilitate transformation of a particular host cell. "Expression cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that allow for enhanced expression of that gene in a foreign host.

The term "expression" refers to the transcription and translation to gene product from a gene coding for the sequence of the gene product. In the expression, a DNA chain coding for the sequence of gene product is first transcribed to a complimentary RNA which is often a messenger RNA and, then, the transcribed messenger RNA is translated into the above-mentioned gene product if the gene product is a protein.

The terms "restriction endonuclease" and "restriction enzyme" refer to an enzyme which binds and cuts within a specific nucleotide sequence within double-stranded DNA.

"Polymerase Chain Reaction" and "PCR" refer to a method that results in the linear or logarithmic amplification of nucleic acid molecules. PCR generally requires a replication composition consisting of, for example, nucleotide triphosphates, two primers with appropriate sequences, DNA or RNA polymerase and proteins. These reagents and details describing procedures for their use in amplifying nucleic acids are provided in U.S. 4,683,202 (1987, Mullis et al.) and U.S. 4,683,195 (1986, Mullis et al.).

The term "sequence analysis software" refers to any computer algorithm or software program that is useful for the analysis of nucleotide or amino acid sequences. "Sequence analysis software" may be

commercially available or independently developed. Typical sequence analysis software will include but is not limited to the GCG suite of programs (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI), BLASTP, BLASTN, BLASTX (Altschul et al., *J. Mol. Biol.* 215:403-410 (1990), and DNASTAR (DNASTAR, Inc. 1228 S. Park St. Madison, WI 53715 USA), and the FASTA program incorporating the Smith-Waterman algorithm (W. R. Pearson, *Comput. Methods Genome Res.*, [Proc. Int. Symp.] (1994), Meeting Date 1992, 111-20. Editor(s): Suhai, Sandor. Publisher: Plenum, New York, NY). Within the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be based on the "default values" of the program referenced, unless otherwise specified. As used herein "default values" will mean any set of values or parameters which originally load with the software when first initialized.

The term "carbon source" refers to a substrate suitable for bacterial cell growth that is distinct from the aromatic substrate. Suitable carbon substrates include but are not limited to glucose, succinate, lactate, acetate, ethanol, monosaccharides, oligosaccharides, polysaccharides, or mixtures thereof.

The term "suicide vector" refers to a vector generally containing a foreign DNA fragment to be expressed in a suitable host cell, coupled with a genetic element that will be lethal to the host cell unless the cell is able to express the foreign DNA. "Suicide vector" is also understood to mean a non-replicating vector capable of transfecting a host cell and facilitating the incorporation of foreign DNA into the genome of the host cell. Such a vector does not replicate and is thus destroyed after incorporation of the heterologous DNA. Examples of common suicide vectors and their construction may be found in Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, 1989.

The meaning of abbreviations is as follows: "h" means hour(s), "min" means minute(s), "sec" means second(s), "d" means day(s), "μL" means microliteres, "mL" means milliliters and "L" means liters.

The nucleic acid fragments of the instant invention may be used to isolate genes encoding homologous proteins from the same or other microbial species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic

acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g. polymerase chain reaction (PCR) , Mullis et al., U.S. Patent No. 4,683,202), ligase chain reaction (LCR), Tabor, S. et al., *Proc. Acad. Sci.* USA 82, 1074, (1985)) or strand displacement amplification (SDA, Walker, et al., *Proc. Natl. Acad. Sci. U.S.A.*, 89, 392, (1992)).

For example, genes encoding similar proteins or polypetides to those of the instant invention could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired bacteria using methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primers DNA labeling, nick translation, or end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part of or full-length of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length DNA fragments under conditions of appropriate stringency.

Typically, in PCR-type amplification techniques, the primers have different sequences and are not complementary to each other. Depending on the desired test conditions, the sequences of the primers should be designed to provide for both efficient and faithful replication of the target nucleic acid. Methods of PCR primer design are common and well known in the art. (Thein and Wallace, "The use of oligonucleotide as specific hybridization probes in the Diagnosis of Genetic Disorders", in *Human Genetic Diseases: A Practical Approach*, K. E. Davis Ed., (1986) pp. 33-50 IRL Press, Herndon, Virginia); Rychlik, W. (1993) In White, B. A. (ed.), Methods in Molecular Biology, Vol. 15, pages 31-39, PCR Protocols: Current Methods and Applications. Humania Press, Inc., Totowa, NJ).

Generally two short segments of the instant sequences may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from

the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding microbial genes.

Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al., *PNAS USA* 85:8998 (1988)) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al., *PNAS USA* 86:5673 (1989); Loh et al., *Science* 243:217 (1989)).

Alternatively the instant sequences may be used as hybridization reagents for the identification of homologs. The basic components of a nucleic acid hybridization test include a probe, a sample suspected of containing the gene or gene fragment of interest, and a specific hybridization method. Probes of the present invention are typically single stranded nucleic acid sequences which are complementary to the nucleic acid sequences to be detected. Probes are "hybridizable" to the nucleic acid sequence to be detected. The probe length can vary from 5 bases to tens of thousands of bases, and will depend upon the specific test to be done. Typically a probe length of about 15 bases to about 30 bases is suitable. Only part of the probe molecule need be complementary to the nucleic acid sequence to be detected. In addition, the complementarity between the probe and the target sequence need not be perfect. Hybridization does occur between imperfectly complementary molecules with the result that a certain fraction of the bases in the hybridized region are not paired with the proper complementary base.

Hybridization methods are well defined. Typically the probe and sample must be mixed under conditions which will permit nucleic acid hybridization. This involves contacting the probe and sample in the presence of an inorganic or organic salt under the proper concentration and temperature conditions. The probe and sample nucleic acids must be in contact for a long enough time that any possible hybridization between the probe and sample nucleic acid may occur. The concentration of probe or target in the mixture will determine the time necessary for hybridization to occur. The higher the probe or target concentration the shorter the hybridization incubation time needed. Optionally a chaotropic agent may

be added. The chaotropic agent stabilizes nucleic acids by inhibiting nuclease activity. Furthermore, the chaotropic agent allows sensitive and stringent hybridization of short oligonucleotide probes at room temperature [Van Ness and Chen (1991) *Nucl. Acids Res.* 19:5143-5151]. Suitable chaotropic agents include guanidinium chloride, guanidinium thiocyanate, sodium thiocyanate, lithium tetrachloroacetate, sodium perchlorate, rubidium tetrachloroacetate, potassium iodide, and cesium trifluoroacetate, among others. Typically, the chaotropic agent will be present at a final concentration of about 3M. If desired, one can add formamide to the hybridization mixture, typically 30-50% (v/v).

Various hybridization solutions can be employed. Typically, these comprise from about 20 to 60% volume, preferably 30%, of a polar organic solvent. A common hybridization solution employs about 30-50% v/v formamide, about 0.15 to 1M sodium chloride, about 0.05 to 0.1M buffers, such as sodium citrate, Tris-HCl, PIPES or HEPES (pH range about 6-9), about 0.05 to 0.2% detergent, such as sodium dodecylsulfate, or between 0.5-20 mM EDTA, FICOLL (Pharmacia Inc.) (about 300-500 kilodaltons), polyvinylpyrrolidone (about 250-500 kdal), and serum albumin. Also included in the typical hybridization solution will be unlabeled carrier nucleic acids from about 0.1 to 5 mg/mL, fragmented nucleic DNA, e.g., calf thymus or salmon sperm DNA, or yeast RNA, and optionally from about 0.5 to 2% wt./vol. glycine. Other additives may also be included, such as volume exclusion agents which include a variety of polar water-soluble or swellable agents, such as polyethylene glycol, anionic polymers such as polyacrylate or polymethylacrylate, and anionic saccharidic polymers, such as dextran sulfate.

Nucleic acid hybridization is adaptable to a variety of assay formats. One of the most suitable is the sandwich assay format. The sandwich assay is particularly adaptable to hybridization under non-denaturing conditions. A primary component of a sandwich-type assay is a solid support. The solid support has adsorbed to it or covalently coupled to it immobilized nucleic acid probe that is unlabeled and complementary to one portion of the sequence.

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening DNA expression libraries. Synthetic peptides representing portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for

peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen DNA expression libraries to isolate full-length DNA clones of interest (Lerner, R. A. *Adv. Immunol.* 36:1 (1984); Maniatis).

The genes and gene products of the instant sequences may be produced in heterologous host cells, particularly in the cells of microbial hosts. Expression in recombinant microbial hosts may be useful for the expression of various pathway intermediates; for the modulation of pathways already existing in the host for the synthesis of new products heretofore not possible using the host. Additionally the gene products may be useful for conferring higher growth yields of the host or for enabling alternative growth mode to be utilized.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

TMO-Containing Bacterial Strains:

Bacterial cells preferred in the present invention are those that possesses the TMO pathway. Such strains are generally restricted to the genus *Pseudomonas* and include, but are not limited to, *Pseudomonas putida* and *Pseudomonas mendocina*. Strains of *Burkholderia* and *Acinetobacter* are also suitable as host cells.

Strains of *Pseudomonas* containing the TMO pathway are known to oxidize toluene to form intermediates of the tricarboxylic acid cycle. PHBA as well as other intermediates, such as *p*-cresol, *p*-hydroxybenzyl alcohol and *p*-hydroxybenzadehyde, are formed in the upper pathway, which metabolizes toluene to the ring cleavage substrate (Figure 1). In wildtype *Pseudomonas* strains, PHBA is immediately converted to protocatechuate (PCA) as it is formed. The biochemistry of the enzymes involved in the upper pathway have been described for several *Pseudomonas* strains (Romine et al., *supra*).

In addition, the homologous TmoST proteins encoded in *P. mendocina* KR1 that control the induction of the toluene monooxygenase pathway induce high levels of expression from the *tmo* pathway genes, in the presence of toluene, and are also useful tools to mediate expression of the catabolic *tmo* genes and PHBA production in any organism that does not possess these genes

Batch and Continuous Fermentations:

The present process uses a batch method of fermentation. A classical batch fermentation is a closed system where the composition of the medium is set at the beginning of the fermentation and not subjected

to artificial alterations during the fermentation. Thus, at the beginning of the fermentation the medium is inoculated with the desired organism or organisms and fermentation is permitted to occur adding nothing to the system. Typically, however, a batch fermentation is "batch" with respect to the addition of carbon source and attempts are often made at controlling factors such as pH and oxygen concentration. In batch systems the metabolite and biomass compositions of the system change constantly up to the time the fermentation is stopped. Within batch cultures, cells moderate through a static lag phase to a high growth log phase and finally to a stationary phase where growth rate is diminished or halted. If untreated, cells in the stationary phase will eventually die.

A variation on the standard batch system is the fed-batch system. Fed-batch fermentation processes are also suitable in the present invention and comprise a typical batch system with the exception that the substrate is added in increments as the fermentation progresses. Fed-batch systems are useful when catabolite repression is apt to inhibit the metabolism of the cells and where it is desirable to have limited amounts of substrate in the medium. An advantage of the fed-batch system is that it is more amenable to the use of toxic or immiscible aromatic substrates such as toluene or *p*-cresol. Using a fed-batch system it is possible to maintain a steady concentration of substrate at non-toxic levels while accommodating maximum bioconversion of the substrate to product.

The production of PHBA from aromatic compounds such as toluene or *p*-cresol will be limited by the amount of the aromatic substrate and carbon sources added. In simple batch fermentation, production will be limited by the amount of toluene initially added. Since toluene is toxic and has limited solubility in water, its low initial concentration will govern the amount of PHBA produced. The ability to run the process at such a low toluene (i.e., 30-60 ppm) allows operation below a lower explosive limit which for toluene is 120 ppm. This low limit is a clear safety advantage to the process. Fed-batch techniques where the carbon source and toluene are added at rates which are similar to the utilization of these compounds will keep the toluene concentration in the medium low and can significantly increase the amount of PHBA produced.

Batch and fed-batch fermentations are common and well known in the art and examples may be found in, for example Brock, Thomas D. In *Biotechnology: A Textbook of Industrial Microbiology*, 2nd ed.; Sinauer

Associates, Inc.: Sunderland, MA, 1989 or Deshpande, Mukund V. *Appl. Biochem. Biotechnol.* 36:227 (1992).

Although the present invention is performed in batch mode, it is contemplated that the method would be adaptable to continuous fermentation methods. Continuous fermentation is an open system where a defined fermentation medium is added continuously to a bioreactor and an equal amount of conditioned medium is removed simultaneously for processing. Continuous fermentation generally maintains the cultures at a constant high density where cells are primarily in log phase growth.

Continuous fermentation allows for the modulation of one factor or any number of factors that affect cell growth or end product concentration. For example, one method will maintain a limiting nutrient such as the carbon source or nitrogen source at low concentration and allow all other parameters to be in excess. In other systems a number of factors affecting growth can be altered continuously while the cell concentration, measured by medium turbidity, is kept constant. Continuous systems strive to maintain steady state growth conditions and thus the cell loss due to medium being drawn off must be balanced against the cell growth rate in the fermentation. Methods of modulating nutrients and growth factors for continuous fermentation processes as well as techniques for maximizing the rate of product formation are well known in the art of industrial microbiology and a variety of methods are detailed by Brock, *supra*.

It is contemplated that the present invention may be practiced using either batch, fed-batch or continuous processes and that any known mode of fermentation would be suitable. Additionally, it is contemplated that cells may be immobilized on a substrate as whole cell catalysts and subjected to fermentation conditions for PHBA production.

Carbon Source:

A variety of carbon sources are suitable in the present invention and include but are not limited to materials (such as succinate, lactate, acetate, ethanol), monosaccharides (such as glucose and fructose), oligosaccharides (such as lactose or sucrose), polysaccharides (such as starch or cellulose), or mixtures thereof and unpurified mixtures from renewable feedstocks (such as cheese whey permeate, cornsteep liquor, sugar beet molasses, and barley malt). The needs of the desired production cell dictate the choice of the carbon substrate. For the purposes of the present invention, glucose is preferred.

Aromatic substrates:

A variety of aromatic substrates may be used in the present invention, including but not limited to toluene, *p*-cresol, *p*-hydroxybenzyl alcohol, *p*-hydroxybenzaldehyde, and any aromatic compounds where the chemical structure is similar to toluene and the intermediates of the TMO pathway (i.e., compounds that are subject to degradation by the TMO pathway).

The concentration of the aromatic substrate (such as toluene and *p*-cresol) and of the carbon source in the medium are limiting factors for the production of PHBA. Preferred concentrations of toluene are from about 30 ppm to about 500 ppm where a range of about 30 ppm to about 60 ppm is most preferred. There are tolerant strains that can ferment toluene at >500 ppm and there are sensitive strains that may operate at a more suitable range of 1-5 ppm. The preferred concentration of *p*-cresol for *Pseudomonas mendocina* is from about 1 mM to about 5 mM. More tolerant strains are expected as well as more sensitive strains. The *p*-cresol concentration needs to be adjusted accordingly.

EXAMPLES

The present invention is further defined in the following Examples. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

Procedures for the genetic manipulations of cellular genomes are well known in the art. Techniques suitable for use in the following examples may be found in Sambrook, J. In *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, 1989.

Materials and methods suitable for the maintenance and growth of bacterial cultures are well known in the art. Techniques suitable for use in the following examples may be found in *Manual of Methods for General Bacteriology*; Phillipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg and G. Briggs Phillips, Eds.; American Society for Microbiology: Washington, DC, 1994 or Brock, Thomas D. In *Biotechnology: A Textbook of Industrial Microbiology*, 2nd ed.; Sinauer Associates, Inc.: Sunderland, MA, 1989. All reagents and

materials used for the growth and maintenance of bacterial cells were obtained from Aldrich Chemicals (Milwaukee, WI), DIFCO Laboratories (Detroit, MI), GIBCO/BRL (Gaithersburg, MD), or Sigma Chemical Company (St. Louis, MO) unless otherwise specified.

Materials and Growth Conditions

Cell strains and plasmids:

For General Use:

Escherichia coli DH5α (Clontech, Palo Alto, CA), *Escherichia coli* DH10B (Gibco BRL, Gaithersburg, MD), *Escherichia coli* JM105 (ATCC 47016), *Escherichia coli* Top10F' (Invitrogen, Carlsbad, CA 92008), *Escherichia coli* XL1-Blue MR (Stratagene, La Jolla, CA), *Escherichia coli* XL2 Blue (Stratagene, La Jolla, CA) and *Escherichia coli* HB101 (Boyer and Roulland-Dussoix, *J. Mol. Biol.* 41(3): 459-472 (1969)).

Sources of DNA for Cloning:

Pseudomonas mendocina KR-1 (U.S. Patent No. 5,171,684; Amgen, Thousand Oaks, CA), *Pseudomonas mendocina* KRC16KDpobA51 (ATCC 55885) (PCT/US98/WO 12072; DuPont, Wilmington, DE), *Pseudomonas putida* (NCIMB 9869) and *Pseudomonas putida* KT2440 (ATCC 47054).

For Plasmid Mobilization:

Escherichia coli S17-1 (ATCC 47055), *Escherichia coli* HB101 (pRK600) (de Lorenzo & Timmis, *Methods Enzymol.* 235: 236-405 (1994)), and *Escherichia coli* CC118 *λpir* (Herrero et al., *J. Bacteriol*, 172(11): 6557-6567 (1990)).

For *pcuC*::*lacZ* Expression:

Escherichia coli MC1061 (CGSC 6649).

For *tmo*::*lacZ* Expression:

Pseudomonas putida KT2440 (Franklin et al., *Proc. Natl. Acad. Sci. USA*, 78(12): 7458-7462 (1981)).

For *p*-Cresol Production:

Escherichia coli G1724 (Invitrogen, Carlsbad, CA), *Escherichia coli* JM105 (ATCC 47016) and *Pseudomonas putida* (ATCC 29607).

For PHBA Production:

Agrobacterium rhizogenes (ATCC 15834), *Pseudomonas mendocina* KRC16KDpobA51 (ATCC 55885) (PCT/US98/12072; DuPont, Wilmington, DE) and *Pseudomonas putida* (ATCC 29607).

For *tmo* and *pcu*-encoded enzyme synthesis:

Pseudomonas putida DOT-T1 (Ramos et al., *J. Bact.* 177(14):3911-3916 (1995)). *Pseudomonas putida* DOT-T1 C5aAR1 has mutations that inactivate toluene dioxygenase. *Pseudomonas putida* DOT-T1E (CECT 5312).

"ATCC" refers to the American Type Culture Collection international depository located at 10801 University Boulevard, Manassas, VA 20110-2209, U.S.A. The designations refer to the accession number of the deposited material.

"CGSC" refers to the *E. coli* Genetic Stock Center located at 355 Osborn Memorial Laboratories, Department of Biology, Yale University, New Haven, CT 06520-8104. The designations refer to the accession number of the deposited material.

"NCCB" refers to the Netherlands Culture Collection of Bacteria, Utrecht University, P.O. Box 80.056, 3508 TB Utrecht, the Netherlands. The designations refer to the accession number of the deposited material.

"NCIMB" refers to the National Collection of Industrial and Marine Bacteria Ltd located at 23 St. Machar Drive, Aberdeen, AB2 1RY, U.K. The designations refer to the accession number of the deposited material.

Growth Conditions:

Typically, studies were conducted by shaking cultures in 125 mL or 250 mL flasks. Experiments using toluene were conducted in 125 mL sealed flasks. Minimal (lean) medium with glucose as the carbon source and ammonia as the nitrogen source was used most extensively. Yeast extract, when added to obtain a "rich" medium, was at 0.5-1.0 g/L. Some of the PHBA production examples included two stages, where the cells were first grown to a suitable cell density in minimal medium containing glucose, followed by transfer to a production medium containing an aromatic substrate for PHBA production. Culture conditions were modulated according to the method of growth and optimized for the production of PHBA. The pH of the cultures should be maintained within a range of about from 6.3 to 7.9. A range of about 7.2 to 7.7 is most preferred. Other media amenable to the procedures of the present invention are common in the art and are fully described in *Manual of Methods for General Bacteriology* (P. Gerhardt, R. G. E. Murray, R. N. Costilow, E. W. Nester, W. A. Wood, N. R. Krieg and G. B. Phillips, Eds.; American Society for Microbiology: Washington, DC, 1994).

EXAMPLE 1

Cloning and Sequencing of the *Pseudomonas mendocina pcu* Operon

Preparation of genomic DNA:

Pseudomonas mendocina KRC16KDpobA51 (ATCC 55885) containing an omega-disrupted *pobA-1* gene was used as the source of genomic DNA. The cells of a 50 mL overnight stationary phase culture were collected by centrifugation at 6,000 rpm, 4 °C for 10 min. The supernatant was decanted and the pellets resuspended with 5 mL TEG (25 mM Tris-HCl, 10 mM EDTA, 50 mM glucose, pH 8.0). About 1.5 mL of RNAse (100 μg/mL) was added into the mixture. The sample was kept at room temperature for 5 min, and then extracted twice with an equal volume of phenol. The two phases were separated by a centrifugation at 6,000 rpm for 10 min. The aqueous phase was extracted twice with phenol:chloroform (1:1). Two volumes of 100% ethanol were added to the aqueous phase to precipitate DNA. After 20 min the solution was centrifuged at 10,000 rpm, and the pellet was collected, dried, and resuspended in 2 to 5 mL TE buffer. The DNA sample was dialyzed against TE buffer at 4 °C overnight.

Construction of a genomic library:

10 μg of genomic DNA was digested with 100 units of BstYI restriction endonuclease at 60 °C, and samples removed at 2, 5, 10, 20 and 30 min intervals in order to obtain partially digested DNA. The pooled partial digests were treated with phenol:chloroform (1:1), chloroform, and two volumes ethanol added to precipitate the DNA. Resuspended DNA (1.6 μg) was ligated at 4 °C overnight using T4 DNA ligase and < 1 μg SuperCos 1 (Stratagene, LaJolla, CA) that had been digested with XbaI, dephosphorylated with CIP, and then digested with BamHI. Each enzyme treatment was followed by extraction with equal volumes of phenol:chloroform (1:1), chloroform, and precipitated with 2 volumes of ethanol. Ligated DNA was recovered in bacteriophage lambda by *in vitro* packaging using a Gigapack II Gold Packaging Extract (Stratagene, La Jolla, CA).

Selection of clones with a *pobA-1* omega insert:

Escherichia coli XL1-Blue MR cells were infected with the packaged cosmid library and plated on LB medium containing 50 mg/L amp and 25 mg/L strep, and cultured at 37 °C overnight. As a control, part of the packaged library was plated on LB medium containing 50 mg/L amp to determine total number of cosmid containing cells. About 1% of the amp

resistant colonies were also strep resistant, and these represented clones that had acquired the omega-inactivated (strR) *pobA-1* gene.

Restriction and hybridization analysis of strR cosmids:

Plasmids were isolated from 5 mL cultures of strR clones using an alkaline lysis method (Birnboim et al., *Nucleic Acids Res.* 7(6):1513-1523 (1979)). The plasmids were digested with the restriction enzymes HindIII or ClaI and fragments separated by electrophoresis overnight on a 0.7% agarose gel in TBE buffer (89 mM Tris-borate, 89 mM boric acid, 2 mM EDTA). Cosmids were identified by the presence of a 14 kb HindIII fragment, or a 12.5 kb ClaI fragment as predicted (Wright et al., *Appl. Environ. Microbiol* 60(1):235-242 (1994)). DNA was transferred from the agarose gel to a GeneScreen Plus nylon membrane (NEN Life Science Products, Boston, MA) using a VacuGene XL system (Pharmacia Biotech, Piscataway, NJ). Depurination of DNA in the gel with 0.25 M HCl for 7 min was followed by denaturation with 1.5 M NaCl + 0.5 M NaOH for 7 min, neutralization with 1.0 M Tris-HCl pH 7.5 + 1.5 M NaCl for 7 min, and transfer to membrane in 20 x SSC for 30 min. The nylon membrane was removed, washed in 0.4 M NaOH (1 min), in 0.2 M Tris-HCl pH 7.5 + 1 x SSC (1 min), in 2 x SSC (1 min), followed by exposure to ultraviolet light for about 2 min to produce nucleic acid crosslinking.

The membrane was prehybridized for 1 h at 65 °C in a hybridization solution containing 5 X SSC, 0.1% (w/v) SDS, 0.5% (w/v) blocking reagent (NEN Life Science Products, Boston, MA) and 5% (w/v) Dextran Sulfate. The hybridization probe was a heterologous sequence for the cytochrome c subunit of PCMH from *Pseudomonas putida* NCIMB 9869. The cytochrome c subunit gene (*pchC*) was cloned from DNA purified from *Pseudomonas putida* NCIMB 9869 by CsCl-ethidium bromide centrifugation (Pemberton et al., *J. Bact.* 114(1):424-433 (1973)), and amplified by PCR using primers (SEQ ID NO:78 and SEQ ID NO:79) based on the published sequence (Kim et al., *J. Bact.* 176(20):6349-6361 (1994)). The 100 μL PCR reaction mixture contained: 0.5 mM dNTPs, reaction buffer (final concentration of 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, and 0.01% gelatin), 0.1 mg of *Pseudomonas putida* genomic DNA, and 1 unit of Taq DNA polymerase. The DNA sample was denatured at 94 °C for 1 min, and annealed at 50 °C for 2 min. Polymerization was performed at 74 °C for 2 min with an increased extention time of 5 sec per cycle. The polymerase chain reaction was accomplished by 25 cycles. The PCR DNA fragment was detected and

analyzed by electrophoresis on 1% agarose gels with 0.5 mg/L ethidium bromide, and cloned into the vector pUC18 (Pharmacia Biotech, Piscataway, NJ).

For ease of identification, the *pchC* DNA was labeled with a fluorescein nucleotide in a 30 μL reaction mixture containing a random primer, reaction buffer, fluorescein nucleotide mix (NEN Life Science Products, Boston, MA) and Klenow enzyme at 37 °C for 1 h. The labeled probe was then hybridized to the membrane-bound genomic DNA in the same buffer for 16 h at 65 °C.

After hybridization, the membrane was washed for 15 min in 2 X SSC, 0.1% SDS, followed by a second 15 min wash in 0.2 X SSC, 0.1% SDS at 65 °C. The membrane was blocked for 1 h in buffer containing 0.5% blocking reagent and then incubated with antifluorescein-horse radish peroxidase conjugate (1:1000) (NEN Life Science Products, Boston, MA) at room temperature for 1 h.

After the incubation the membranes were washed four times for 5 min with 0.1 M Tris-HCl pH 7.5, 0.15 M NaCl, and incubated in a chemiluminescence reagent (Renaissance nucleic acid chemiluminescent reagent, NEN Life Science Products, Boston, MA) for 1 min at room temperature, and then exposed to Reflection autoradiography film (NEN Life Science Products, Boston, MA). Those clones having both the correct restriction pattern with HindIII or ClaI, and which hybridized to the *pchC* probe, were selected for sub-cloning and sequencing.

Subcloning and sequencing:

A strR cosmid was digested with HindIII and the ~14 kb insert isolated from a 0.8% agarose gel using the DNA preparation kit GeneClean (Bio101, Vista, CA). The isolated fragment was cloned into the HindIII site of the vector pZErO-1 (Invitrogen, Carlsbad, CA), transformed into *Escherichia coli* Top10F', and selected on LB medium containing 50 mg/L zeocin. Zeocin-resistant clones were screened by digestion of plasmid minipreps with HindIII, BamHI, SalI/BamHI, ClaI/SphI, and SphI. A plasmid with a digestion pattern indicating that the *pcu*-encoded enzymes were oriented for transcription by the *lac* promoter of pZErO-1 was designated pPCU1, and a plasmid with the opposite orientation was designated pPCU2.

A 3.5 kb NruI/EcoRI fragment was isolated from pPCU1, and a BamHI adaptor (New England Biolabs, Beverly, MA) annealed and ligated to 2 μg of fragment in a 20 μL reaction containing 2 mM adaptor at 16 °C

for 16 h. Following a phenol:chloroform (1:1) extraction and ethanol precipitation, the DNA was dissolved in 12 μL TE, digested with BamHI for 5 h, and purified by electrophoresis on a 1% agarose gel and isolated with GeneClean as before. The BamHI/EcoRI fragment was cloned into the EcoRI/BamHI digested vector pK194 (ATCC 37767) to yield plasmid pPCU3. The complete sequence of the *pcu* operon is shown in SEQ ID NO:1 and the nucleotide sequences for the transcriptional activator PcuR (SEQ ID NO:98), PHBAD (SEQ ID NO:99), the two subunits of PCMH (SEQ ID NO:100 and SEQ ID NO:102), and an unidentified open reading frame (SEQ ID NO:101). Also given are the predicted amino acid sequences for the transcriptional activator PcuR (SEQ ID NO:2), PHBAD (SEQ ID NO:3), the two subunits of PCMH (SEQ ID NO:4 and SEQ ID NO:6), and an unidentified open reading frame (SEQ ID NO:5). The DNA was sequenced with synthetic primers (SEQ ID NO:7 to SEQ ID NO:77) according to standard methods.

Identification of the PHBAD and PCMH coding sequences was based on percent homolgy to the corresponding predicted amino acid sequences for these enzymes from *Pseudomonas putida* NCIMB 9866 and 9869 (Kim et al., *supra*; Cronin et al., *DNA Sequence* 10(1):7-17 (1999)). Identification of the PcuR transcriptional activator was based on homology to the predicted amino acid sequence of the TbuT transcriptional activator of *Ralstonia pickettii* (Olsen et al., *J. Bacteriol.* 176(12):3749-3756 (1994)). Based on the work of Cronin et al. (*supra*), the unidentified open reading frame (SEQ ID NO:5) may be an inner membrane protein. Their analysis by PSORT for the *Pseudomonas putida* protein predicts it to be an inner membrane protein, and analysis by TMpred predicts it to have one or two transmembrane helices, with the bulk of the protein lying on the cytoplasmic side in either situation. The arrangement of genes in the *pcu* operon is illustrated in Figure 2. The best homologies to each ORF, and thus their putative function in the *pcu* operon, are listed in Table 1.

Table 1

ORF	Similarity Identified	SEQ ID base	SEQ ID Peptide	% Identity[a]	% Similarity b	E-value[c]	Citation
1.1	gi|1657782 transcriptional activator TbuT (*Ralstonia pickettii*)	98	2	48%	63%	1e-143	*J. Bacteriol.* 176 (12), 3749-3756 (1994)
1.2	gb|AAA75634.2| *p*-hydroxybenzaldehyde dehydrogenase (*Pseudomonas putida*)	99	3	75%	83%		*DNA Seq.* 10 (1), 7-17 (1999)
1.3	gb|AAA80319.2| *p*-cresol methylhydroxylase, cytochrome subunit precursor (*Pseudomonas putida*)	100	4	60%	74%	3e-25	*DNA Seq.* 10 (1), 7-17 (1999)
1.4	gb|AAD29836.1|U96338_3 unknown (*Pseudomonas putida*)	101	5	46%	61%	5e-36	*DNA Seq.* 10 (1), 7-17 (1999)
1.5	gb|AAA80318.2| *p*-cresol methylhydroxylase, flavoprotein subunit (*Pseudomonas putida*)	102	6	78%	88%		*DNA Seq.* 10 (1), 7-17 (1999)
2.1	emb|CAB43725.1| membrane protein (*Pseudomonas putida*)	103	92	81%	87%		*Gene* 232, 69-76 (1999)
tmoS	emb|CAB43735.1| bZIP histidine kinase (*Pseudomonas putida*)	114	116	83%	89%		*Gene* 232, 69-76 (1999)
tmoT	emb|CAB43736.1| response regulator (*Pseudomonas putida*)	115	117	85%	93%		*Gene* 232, 69-76 (1999)

[a]%Identity is defined as percentage of amino acids that are identical between the two proteins.

[b]% Similarity is defined as percentage of amino acids that are identical or conserved between the two proteins.

[c]Expect value. The Expect value estimates the statistical significance of the match, specifying the number of matches, with a given score, that are expected in a search of a database of this size absolutely by chance.

EXAMPLE 2

Cloning the *Pseudomonas mendocina tmo* Operon

Pseudomonas mendocina KR-1 was the source of total genomic DNA, and it was isolated as described before for *Pseudomonas mendocina* KRC16KDpobA51(ATCC 55885). Total genomic DNA was digested with SstI + XmaI, separated on a 0.8% low-melting agarose gel, and fragments in the 5-7 kb size range recovered. The purified DNA was ligated to the vector pUC18 that had been digested with SstI + XmaI, and the ligated DNA transformed into *Escherichia coli* JM105. Clones were selected on LB plates containing amp (100 mg/L) and 10 mM tryptophan. *Escherichia coli* is able to produce indole from tryptophan using tryptophanase, and the *tmo*-encoded toluene monooxygenase converts the indole to *cis*-indole-2,3-dihydrodiol, which then forms indoxyl through the spontaneous elimination of water, and is then oxidized by air to indigo. An indigo-producing colony was isolated and the correctly configured plasmid identified as pTMO1.

EXAMPLE 3

Construction of *pcu* and *pcu/tmo* Expression Plasmids

Construction of the *pcu* plasmid pPCU12:

pPCU1 was digested with NruI + ApaI and a 2.4 kb fragment was isolated by electrophoresis on a 1% agarose gel and purified using a GeneClean kit, then ligated to the SmaI + ApaI digested vector pGadGH (Clontech, Palo Alto, CA). The ligation was transformed into competent *Escherichia coli* strain DH5α, and transformants were isolated on LB + amp (100 mg/L) plates. The correct construct, which was identified by the band patterns produced with HindIII + BamHI or BamHI + SalI digests, was named pPCU9.5. Next, a 2.6 kb ApaI fragment was isolated from pPCU1 by electrophoresis on a 1% agarose gel followed by purification with GeneClean as before. This fragment was cloned into ApaI-digested pPCU9.5 which had also been treated with CIP. Clones containing the inserted fragment were distinguished by digestion with ApaI and detected the presence of the 2.6 kb fragment. The orientation of the insert was determined by the fragmentation pattern of a BglII digest. The plasmid with the pattern indicating a complete *pcu* operon was named pPCU10.

The ~5 kb BamHI + HindIII fragment from pPCU10 was isolated as before and ligated into the BamHI + HindIII sites in the vector pK184 (ATCC 37766). The ligation was transformed into ultracompetent XL2 Blue cells. Transformants were selected using LB + kan (50 mg/L) plates.

EcoRI and BglII digests were used to determine the correct construct, which was named pPCU11. The 5 kb BamHI + HindIII fragment was isolated from pPCU11 as described above, and the single-stranded ends were converted to double strands with the Klenow fragment of DNA polymerase I. The vector pRK310 (Ditta et al., *Plasmid* 13:149-153 (1985)) was digested with HindIII, and the single-stranded ends were also treated with the Klenow fragment of DNA polymerase 1 and then phosphatased with CIP. The two fragments were ligated together and electroporated into Electromax DH10B cells. Colonies with plasmids were selected on LB + tet (12.5 mg/L) plates. EcoRI and SalI digests of plasmids from the colonies were used to identify a clone of the correct construction, named pPCU12.

Construction of the *pcu* plasmid pPCU18:

A 7.5 kb MluI + NheI pPCU1 fragment was isolated through agarose gel electroporesis followed by purification with GeneClean. It was ligated into the MluI + NheI sites of plasmid pSL1180 (Pharmacia Biotech, Piscataway, NJ). The ligation was transformed into competent DH5α cells. Transformants were identified by growth on LB + amp (100 mg/L) plates. SalI digests indicated the correct construct, which was named pPCU17. Plasmid pPCU17 was digested with BamHI + HindIII, and the 7.5 kb piece of DNA with the *pcu* genes was isolated as described earlier. The fragment was cloned into the BamHI + HindIII sites of the vector pGV1120 (Leemans et al., Gene 19:361-364 (1982)). Electrocompetent *Pseudomonas putida* strain DOT-T1 C5aAR1 cells were electroporated with the ligated DNA. Cells were selected on LB + strep (50 mg/L) plates at 30 °C overnight. Plasmids were isolated from clones grown on the plates and digested with EcoRI. The plasmid with the correct digest pattern was named pPCU18.

Construction of *pcuC*::*lacZ* fusion plasmids pPCUR1 and pPCUR2:

The non-translated *pcu* promoter region between *pcuR* and *pcuC* was amplified by PCR in order to construct a *lacZ* fusion to examine regulation of the *pcu* operon. The reaction contained the following: 0.5 μL pPCU1 (0.8 μg/μL), 1 μL primer PCUR1L (10 pmol/μL) (SEQ ID NO:95), 1 μL primer PCUR2L (10 pmol/μL) (SEQ ID NO:96), 33.3 μL water, 2.2 μL 25 mM $Mg(OAc)_2$, 1 μL 10 mM dNTPs, 10 μL 5X GC Genomic PCR Reaction Buffer, and 1 μL Advantage-GC Genomic Polymerase Mix (50X). The last four components were from the Advantage-GC Genomic PCR Kit (Clontech, Palo Alto, CA). The reaction was put through the following

thermocycles: 1 min at 94 °C, then 30 cycles of 30 sec at 94 °C, 4 min at 68 °C, and incubation at 4 °C overnight. The PCR product was purified using GeneClean, digested with BamHI and isolated as a 2.4 kb fragment following electrophoresis on a 0.6% agarose gel. The fragment was ligated to the vector pMC1403 (NCCB no. PC-V3088), which had been digested with BamHI and dephosphorylated with SAP. The ligation was transformed into competent *Escherichia coli* MC1061 cells. Transformants were selected on LB + Amp (100 mg/L) plates. The orientation of the insert in the vector was determined by SstI and PstI digests, and a plasmid where the ribosome binding site and AUG initiation codon from *pcuC* was fused to the *lac* operon was named pPCUR1. A control plasmid with the PCR product cloned in the opposite orientation was named pPCUR2.

Construction of the *tmo* plasmid pTMO3:

The vector pLEX (Invitrogen, Carlsbad, CA) was digested with SphI + SstI and ligated to a 6 kb *tmo* fragment from pTMO1 (Figure 3) digested with the same enzymes. Ligated DNA was transformed into *Escherichia coli* strain G1724 (Invitrogen, Carlsbad, CA) and selected on LB + amp (100 mg/L). A plasmid with *tmo* under the transcriptional control of the P_L promoter was designated pTMO3.

Construction of the *tmo* plasmid pTMO9:

Plasmid pTMO1 was digested with HindIII + BglII. The 960 bp fragment was isolated and purified with GeneClean, and ligated to HindIII + BglII cut plasmid pSL1180 (Pharmacia, Piscataway, NJ). The ligation was used to transform competent *Escherichia coli* XL2 Blue cells, which were then incubated on LB + amp (100 mg/L) plates. HindIII digests and NcoI digests of the plasmids from transformants identified those with the correct insert. A correct plasmid was named pTMO6. The 960 bp SmaI + HindIII fragment from pTMO6 was isolated and purified as before and ligated to the vector pMMB208 (ATTC 37810) which had been digested with SmaI + HindIII. Competent XL1 Blue cells were transformed with the ligated DNA and spread onto LB + chl (50 mg/L) plates. HindIII + SstI digests of plasmids from transformants were used to determine clones with the proper constructs, which were named pTMO7. Next, a 5 kb piece of DNA was isolated from pTMO1 by BglII + BamHI digestion and inserted into the BamHI site of pTMO7. The ligated DNA was transformed into competent XL1 Blue cells, which were then spread onto LB + chl (50 mg/L) plates and incubated at 37 °C until colonies were apparent. After a few days at 4 °C, some of the colonies on the plates developed an

indigo-blue color. Plasmids were isolated from indigo-blue colonies and digested with HindIII to confirm the presence of a correctly constructed plasmid, which was named pTMO8. The 1.2 kb kan resistance marker from pUC4K (Pharmacia, Piscataway, NJ) was isolated by EcoRI digestion, gel electrophoresis, and GeneClean purification. It was ligated to EcoRI cut and SAP treated pTMO8, then transformed into competent XL1 Blue cells. The correct plasmid from a clone that grew on LB + kan (50 mg/L) + chl (50 mg/L) plates was named pTMO9.

Construction of the *tmo* plasmids pTMO17 and pTMO18:

A BamHI digest of pTMO11 and a BglII digest of the vector pGV1120 (Leemans et al., *Gene* 19:361-364 (1982)) were electrophoresed on a 0.8% agarose gel. The 6 kb pTMO11 fragment and the vector fragment were excised and purified using a GeneClean kit. The two pieces were ligated together, transformed into competent *Esvcherichia coli* DH5α cells, and spread onto LB + tet (10 mg/L) plates. The plasmids from selected colonies were digested with HindIII, and one with the correct pattern of bands was named pTMO17.

A 7.5 kb BamHI pTMO15 fragment and a BglII fragment from the vector pGV1120 were gel-purified as described earlier. They were ligated together, transformed into *Escherichia coli*, and plated on LB + tet (10 mg/L) plates. HindIII digests of the plasmids from transformants were used to identify constructs containing the *tmo* operon, and a correctly configured plasmid was named pTMO18.

Construction of the expression plasmid pMC3 containing *pcu* and *tmo*:

The *pcu* operon (*pcuC* through *pcuB*) was amplified in a PCR reaction containing 4 μL dNTPs (2.5 mM), primer PCUAMP1 (10 pmol/μL) (SEQ ID NO:93), primer PCUAMP2 (10 pmol/μL) (SEQ ID NO:94), 30.7 μL water, 0.3 μL pPCU10 (0.3 μg), 2 μL Buffer A, 8 μL Buffer B, and 1 μL Elongase (the last 3 components were from the Elongase amplification system (Gibco BRL, Gaithersburg, MD). The cycles used were as follows: 30 sec at 94 °C, then 35 cycles of (45 sec at 94 °C, 30 sec at 55 °C, 5 min at 68 °C), finally 4 °C overnight. The ~5.5 kb product was purified using a GeneClean kit, digested with HindIII, isolated from a 0.8% agarose gel, and purified again with GeneClean. This fragment was inserted into a HindIII digested and phosphatased (using CIP) pUC18 vector. The ligation was transformed into competent *Escherichia coli* XL1 Blue cells, and transformants were selected on LB + amp (100 mg/L) + IPTG (1 mM) + X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) (50 mg/L) plates.

White colonies indicated the presence of an insert in the vector. The correct construct, which was named pPCU14, was found by observing HindIII digest patterns of the plasmids isolated from white transformants. Orientation of the inserts was determined by PstI digest patterns.

To remove the BamHI site in vector pRK310, the single-stranded ends created by a BamHI digest of pRK310 were removed using mung bean nuclease. The vector was then allowed to self-ligate and the product was electroporated into *Escherichia coli* Electromax DH10B cells. The cells were spread onto LB + tet (12.5 mg/L) plates to select for those containing plasmids. BamHI + BglII digests were used to identify clones that had the correct construct, which was named pRK310BamKO. This new vector was digested with HindIII and phosphatased with SAP. The 5.5 kb HindIII fragment of pPCU14, isolated as described previously, was ligated into the vector pRK310BamKO. Electromax DH10B cells were electroporated with the ligated DNA. Plasmids were isolated from cells that grew on LB + tet (12.5 mg/L) plates and were digested, first with HindIII to ascertain the presence of an insert, then with SalI to determine the orientation of the insert. The correct plasmid was named pPCU16.

The 1.2 kb kan resistance marker from the vector pUC4K (Pharmacia Biotech, Piscataway, NJ) was isolated as an EcoRI fragment in the manner described above, and inserted into the EcoRI site of pTMO1. The ligation was transformed into competent XL1 Blue cells, which were then spread onto LB + kan (50 mg/L) + amp (100 mg/L) + IPTG (1 mM) plates. Indigo-blue colonies were diagnostic for the presence of the *tmo* operon because *tmo*-encoded toluene monooxygenase catalyzes, in part, the formation of indigo from indole. The correct construct, which was named pTMO11, was ascertained through the digestion of the transforming plasmids with BamHI. The 5.9 kb pTMO11 fragment containing the *tmo* genes was purified as described previously, and was ligated to BamHI cut and SAP-phosphatased pPCU16. The ligated DNA was electroporated into Electromax DH10B cells, which were then spread onto LB + tet (12.5 mg/L) + IPTG (1 mM) plates. Transformants that carried plasmids with the *tmo* genes were indigo-blue, as described before. The correct construct was identified by digestion with PstI, and was named pMC3.

Construction of the expression plasmid pMC4 containing *pcu* and *tmo*:

Pseudomonas mendocina KR-1 genomic DNA was digested to completion with EcoRI. The digested DNA was run on a 0.8% agarose

gel, and DNA larger than ~6 kb was cut out of the gel and purified with GeneClean. Plasmid pUC18 was digested with EcoRI and the ends were phosphatased with SAP. The genomic DNA pieces were ligated to the vector, then electroporated into *Escherichia coli* Electromax DH10B cells. The cells were incubated on LB + amp (100 mg/L) + IPTG (1 mM) plates. Plasmids were isolated from indigo-producing transformant colonies and digested with EcoRI. The plasmid with the correct digest pattern was named pTMO14. A 7.3 kb SmaI fragment from pTMO14 was isolated as before and cloned into the 2.7 kb HincII cut and SAP treated pUC4K vector. The ligation was used to electroporate electrocompetent *Escherichia coli* DH5α cells, which were then incubated on LB + amp (100 mg/L) + IPTG (1 mM) plates. BamHI digests were performed on plasmids from indigo-blue colonies from the plates. The correct construct, which had the *tmo* operon flanked by BamHI sites, was named pTMO15.

The 7.3 kb BamHI pTMO15 fragment was isolated as before and inserted into the BamHI site of pPCU18. This ligated DNA was electroporated into electrocompetent *Pseudomonas putida* DOT-T1 C5aAR1 cells, which were incubated on LB + strep (50 mg/L) + indole (1 mM) plates. Some of the plates also had a drop of toluene added to the inside of the top lid. They were all incubated at 30 °C overnight. PstI digests of plasmids from transformants identified one that had both *pcu* and *tmo* operons, and this clone was named pMC4 (Figure 4).

EXAMPLE 4

Production of *p*-Cresol from Toluene in *Escherichia coli*

Escherichia coli strain JM105 harboring plasmid pTMO1, which places *tmo* expression under control of the *lac* promoter, was grown under inducing conditions in the presence of 1 mM IPTG, or under non-inducing conditions in the absence of IPTG. *Escherichia coli* strain G1724 harboring plasmid pTMO3, which places *tmo* expression under control of the P_L promoter, was grown under inducing conditions in the presence of 100 mg/L tryptophan, or under non-inducing conditions in the absence of tryptophan.

Induced or non-induced cell samples were resuspended in minimal medium at a concentration of 100 mg/mL. To 26 mL of minimal medium in a 125 mL sealed flask was added 4 mL of the cell suspension, and 1 mL of toluene placed in a center well. Following a 36 h incubation 15 mL of the cells were acidified, extracted with ethyl acetate, and analysed by GC/MS. Table 2 shows that *p*-cresol is produced when induced cells

harboring either plasmid pTMO1 or plasmid pTMO3 are incubated in the presence of toluene. In contrast, in the absence of induction of *tmo* using either plasmid, no *p*-cresol is detectable.

TABLE 2

Plasmid	Inducer	GC Peak Area
pTMO1	IPTG	1.45×10^6
pTMO1	None	0
pTMO3	Tryptophan	5.05×10^5
pTMO3	None	0

EXAMPLE 5

Bioconversion of Toluene to *p*-cresol in *Pseudomonas putida* ATCC 29607

Pseudomonas putida ATCC 29607 was transformed with pTMO9 and pPCU12, grown at 30 °C and 250 rpm in medium A (Table 3). At an OD_{600} of 1.98 (16 h) cells were harvested and washed in MM#4 medium (Table 4). (Trace elements found in both medium A and MM#4 Medium are listed in Table 5.) PHBA production was carried out in 125 mL sealed flasks in 5 mL MM#4 medium that contained 0.5 OD_{600} cells, 0.05 mM $MgSO_4$, 2 mM glucose, 1 mM IPTG, 0.1 M HEPES buffer pH 7.5-8.0 and 60 ppm toluene. The flasks were incubated shaking at 250 rpm and 30 °C. A non-induced control did not have IPTG added. Samples were incubated for 6 h, and the presence of *p*-cresol detected by HPLC. In the presence of IPTG 0.93 mM *p*-cresol was present after 6 h, compared to 0.135 mM in the non-induced sample.

Table 3

Medium A

	per L	Special Conditions
KH_2PO_4	1.2 g	
$(NH4)_2SO_4$	3 g	
glucose	7 g/L	sterilized separately
$MgSO_4 \cdot 7H_20$	0.15 g	
trace elements	10 mL	sterilized separately
HEPES	0.05 M	
yeast extract	1 g	sterilized separately
Titrate to pH 7.2 with KOH or H_2SO_4		

TABLE 4
MM #4 Medium

trace elements	10 mL
yeast extract	0.48 g
$MgSO_4 \cdot 7H_20$	10 mM
$NaKPO_4$	25 mM
DD H_2O	1 L
PH	7.2

TABLE 5
Trace Elements in Medium A and MM#4

	g/L
citric acid	10
$CaCl_2 \cdot 2H_2O$	1.5
$FeSO_4 \cdot 7H_2O$	2.8
$ZnSO_4 \cdot 7H_2O$	0.39
trace elements	
Medium A and MM#4	g/L
$CuSO_4 \cdot 5H_2O$	0.38
$CoCl_2 \cdot 6H_2O$	0.2
$MnCl_2 \cdot 4H_2O$	0.3

EXAMPLE 6

Identification and Sequence of a *tmo* Regulatory Region

Detection of a regulatory sequence:

Plasmids pTMO17 and pTMO18 differ in the amount of *tmo* sequence information that is present. Plasmid pTMO17 contains the six toluene monooxygenase genes *tmoA-F*. Plasmid pTMO18 also contains *tmoA-F*, but in addition has 1326 bp of DNA sequence information upstream from the translational initiation codon of *tmoA*. Plasmids pTMO17 and pTMO18 were transformed separately into *Pseudomonas putida* DOT-T1 C5aAR1 and selected on LB + strep (100 mg/L). Colonies were inoculated into 25 mL LB + 1 mM indole + strep (100 mg/L) and shaken in a 125 mL baffel flask at 200 rpm and 30 °C until indigo production occured. A 5 mL sample of cell suspension was extracted twice with an equal volume of ethyl acetate to solubilize the indigo, the two extracts were combined and the absorption at 600 nm recorded. A

standard curve prepared with pure indigo in ethyl acetate was used to determine amounts in cell extracts.

TMO enzyme assays were carried out in a separate experiment using the same plasmids and strain. TMO was measured spectrophotometrically using a coupled assay, linking phenazine ethosulfate (PES) oxidation to reduction of 2,6-dichlorophenol-indophenol (DCPIP) as measured by a decrease in absorption at 600 nm (E_{600nm} = 21,000 $M^{-1}cm^{-1}$). The assay was initiated by the addition of enzyme to a 2.0 mL reaction mixture containing 0.67 μmol PES, 0.1 μmol DCPIP, 1.0 μmol toluene and saturating levels of purified *p*-cresol methylhydroxylase.

Table 6 shows that the presence of additional DNA upstream of *tmoA* enhances the level of TMO activity, which leads to a considerable improvement in indigo prduction.

TABLE 6

Plasmid	TMO activity	Indigo produced (mg/L)
pTMO17	0.7	2.5
pTMO18	1.3	88.0

Sequence of *tmoX* and its upstream promoter region:

The DNA upstream of *tmoA* was sequenced with synthetic primers (SEQ ID NO:80 to SEQ ID NO:90) according to standard methods. The complete sequence of the DNA has the sequence found in SEQ ID NO:91. Encoded within the sequence is a protein, TmoX, with the initiator methionine at nucleotide 192 and a TAA translation terminator at position 1560. The predicted amino acid sequence of TmoX is given as SEQ ID NO:92 and its nucleotide sequence is in SEQ ID NO:103. TmoX has an 81% identity (87% similarity) in its predicted amino acid sequence compared to that of the TodX protein of *Pseudomonas putida* DOT-T1 (Table 1). TodX has been described as an outer membrane protein that may be involved in facilitating the delivery of exogenous toluene inside cells (Wang et al., *Mol. Gen. Genet.* 246:570-579 (1995)), but has also been linked to the signal transduction process which results in specific response of a *tod* promoter to toluene (Lau et al., *Proc. Natl. Acad. Sci. USA* 94:1453-1458 (1997)).

The *tmoX* promoter was identified by primer extension using a 23-mer oligonucleotide (SEQ ID NO:97) complementary to the DNA coding strand. The first nucleotide of the primer corresponded to a

nucleotide 200 bp downstream from the A of the ATG initiation codon of the *tmoX* gene. *Pseudomonas mendocina* KR-1 was grown overnight in M9 minimal medium with 10 mM citrate as the sole carbon source. To 200 mL of fresh medium was added 5 mL of overnight culture to give an initial OD of about 0.2 at 660 nm. The culture was incubated at 30 °C on a rotary shaker to an OD of 0.8 at 660 nm. Aliquots of 20 mL were supplemented with either 1 mM *p*-cresol, toluene in the gas phase, or a control with no additions. Samples were used for RNA isolation at 30, 60 and 180 min after addition of the effector.

The primer (SEQ ID NO:97) was labeled at its 5' end using ^{32}P-γ-ATP and polynucleotide kinase. To 30 μg of total RNA for each sample were added 10^5 CPM of labeled primer, which was extended using reverse transcriptase. The resulting cDNA was separated on a urea-polyacrylamide sequencing gel. In addition, the labeled primer was used to establish a sequencing ladder to facilitate the identification of the transcription initiation point. It was established that the 260 base cDNA product positioned the *tmoX* transcription initiation point as a G located 60 bp upstream of the A of the ATG translation initiation codon of *tmoX*. Analysis of the region upstream of the ATG codon shows the presence of a prokaryotic Shine-Dalgarno sequence. Also noted is the presence of -10 and -35 sequences upstream of the transcript initiation site, each positioned respectively at bp 124-128 and bp 101-105 in SEQ ID NO:91. A putative TodT motif is to be found at bp 30-46 in SEQ ID NO:91.

By comparing the amounts of cDNA obtained under different induction regimes, it was found that growth in the presence of toluene led to a 20-fold increase in *tmoX* mRNA compared to growth on citrate, with a maximal level observed 30 min after exposure to the solvent. The transcription of *tmoX* was also induced by the presence of *p*-cresol, with maximal levels also at 30 min, followed by a decrease in signal intensity probably related to exhaustion of the inducer in the culture medium.

EXAMPLE 7

Regulation of *pcu* Expression by PCUR

The *pcuC::lacZ* fusion plasmid pPCUR1, and the control plasmid pPCUR2, were transformed into *Escherichia coli* MC1061. Plasmids pPCUR1 and pPCUR2 also encode PcuR and amp resistance. Cultures were grown overnight in flasks shaking at 37 °C in M9 minimal medium containing 1% glucose and 50 mg/L amp. In addition, some flasks also contained intermediate compounds of the toluene to PHBA pathway,

including toluene and PHBA. The following were added at a concentration of 1 mM to separate flasks prior to overnight incubation: *p*-cresol, *p*-hydroxybenzylalcohol, *p*-hydroxybenzaldehyde and PHBA. Toluene (5 μL) was added to the gas phase of a 125 mL sealed flask. The overnight cultures were treated with chloroform and SDS, and assayed for β-galactosidase as described in J. H. Miller in *A Short Course in Bacterial Genetics* (Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY; 1992).

Table 7 shows that when using plasmid pPCUR1 there is no induction of the *pcuC::lacZ* gene fusion when toluene or PHBA are present, neither of which are substrates for enzymes encoded by the *pcu* operon. In contrast, the presence of *p*-cresol, *p*-hydroxybenzylalcohol or *p*-hydroxybenzaldehyde all lead to significant induction of *pcuC::lacZ*, and all three compounds are substrates for the two enzymes encoded by the *pcu* operon i.e. PCMH and PHBAD. In the control plasmid pPCUR2, in which the *pcuC* gene is incorrectly orientated for expression, the presence of *p*-cresol does not lead to expression of β-galactosidase activity.

TABLE 7

Plasmid	Inducer	β-galactosidase units
pPCUR1	toluene	0.55
pPCUR1	*p*-cresol	19.53
pPCUR2	*p*-cresol	0.05
pPCUR1	*p*-hydroxybenzylalcohol	5.65
pPCUR1	*p*-hydroxybenzaldehyde	9.70
pPCUR1	PHBA	0.06

EXAMPLE 8

Activity of Plasmid-Encoded Enzymes in *Pseudomonas putida* DOT-T1

Cells were grown in Medium A with the appropriate antibiotic in shake flasks at 30 °C (200 rpm). When the glucose had been depleted, the induction phase was initiated by addition of toluene and/or *p*-cresol. Three consecutive additions of inducer were made, each separated by one hour. For induction with IPTG, the compound was added at a concentration of 1 mM. Cells were collected by centrifugation, washed once with phosphate buffered saline and stored at -80 °C until assay.

TMO was measured spectrophotometrically using a coupled assay, linking phenazine ethosulfate (PES) oxidation to reduction of 2,6-dichlorophenol-indophenol (DCPIP) as measured by a decrease in

absorption at 600 nm (E_{600nm} = 21,000 $M^{-1}cm^{-1}$). The assay was initiated by the addition of enzyme to a 2.0 mL reaction mixture containing 0.67 μmol PES, 0.1 μmol DCPIP, 1.0 μmol toluene and saturating levels of purified *p*-cresol methylhydroxylase (PCMH).

p-Cresol methylhydroxylase (PCMH) activity was measured spectrophotometrically using a coupled assay, linking phenazine ethosulfate (PES) oxidation to reduction of 2,6-dichlorophenol-indophenol (DCPIP) as measured by a decrease in absorption at 600 nm (E_{600nm} = 21,000 $M^{-1}cm^{-1}$). The assay was initiated by the addition of enzyme to a 2.0 mL reaction mixture containing 0.67 μmol PES, 0.1 μmol DCPIP and 1.0 μmol *p*-cresol. This assay was also used to measure toluene monooxygenase (TMO) activity by substituting 0.5-1.0 μmol toluene into the reaction mixture and by the addition of saturating levels of purified *p*-cresol methylhydroxylase (PCMH).

p-Hydroxybenzoate dehydrogenase (PHBAD) activity was measured spectrophoto-metrically using a reaction mix containing 600 nmol $NADP^+$, 40 nmol *p*-hydroxybenzaldehyde, 1.0 mL of 50 mM glycine-NaOH (pH 9.6), and an appropriate amount of enzyme. Enzyme activity was determined by an increase in absorbance at 330 nm. A unit of activity is the amount of enzyme required to oxidize 1.0 μmol of *p*-hydroxybenzaldehyde per min ($E_{330\ nm}$ = 28,800 $M^{-1}cm^{-1}$).

p-Hydroxybenzoate hydroxylase (PHBH) was assayed spectrophotometrically by following the oxidation of NADPH. The reaction mixture contained 250 nmol NADPH, 700 nmol *p*-hydroxybenzoate, an appropriate amounts of enzyme, and 50 mM Tris-HCl buffer (pH 8.0) to give a final volume of 1.0 mL. A unit of activity is the amount of enzyme required to oxidize 1.0 μmol of NADPH per min ($E_{340\ nm}$ = 6,200 $M^{-1}cm^{-1}$).

Enzyme assays for PCMH and PHBAD demonstrate that both of the *Pseudomonas mendocina pcu* enzymes are expressed in *Pseudomonas putida* strain DOT-T1 (Table 8). In addition, it is noteworthy that expression of *pcu* is superior when using its native promoter in plasmid pMC4 compared to the use of a *lac* promoter in plasmid pMC3. This is also true for TMO, where greater activity is seen when using the endogenous *tmo* promoter in pMC4 when compared to the *lac* promoter in plasmid pMC3.

TABLE 8

Plasmid	Promoter	Inducer	TMO	PCMH	PHBAD
pMC3	*lac*	IPTG	0.76	0.74	
pMC4	*pcu* or *tmo*	Toluene	18.2	13.5	9.0
pPCU18	*pcu*	*p*-cresol	0	16.1	3.6
none	-	-	0	0.29	0.05

EXAMPLE 9

Production of PHBA from *p*-Cresol by *Pseudomonas putida* ATCC 29607 Transformed with a *pcu* Expression Plasmid

The mobilizing *Escherichia coli* strain S17-1 was used to introduce the *pcu* expression plasmid pPCU12 into *Pseudomonas putida* ATCC 29607 by conjugation. A single colony of S17-1 having the plasmid pPCU12 was inoculated in 20 mL LB medium and grown at 37 °C to log phase. Another colony of *Pseudomonas putida* ATCC 29607 was inoculated in 20 mL LB medium and incubated at 30 °C and grown to log phase. The cells of both cultures were washed twice with LB medium and resuspended in LB medium. S17-1 cells harboring pPCU12 and *Pseudomonas putida* ATCC 29607 were mixed at a ratio of 1:4 and were plated on agar plates of LB medium. The plates were incubated at 30 °C for 8 h. The cells were collected and then plated on agar plates containing phosphate buffer, 1 mM succinate, 10 mM strep and 25 mg/L kan, and kan resistant colonies were selected. Transformants, or a non-transformed control strain, were grown in 15 mL M9 minimal medium containing 1% glucose, 5 mM *p*-cresol, 10 mM $MgSO_4$, tet (15 mg/L) in 125 mL flasks at 30 °C and 225 rpm. Samples were removed at the indicated timepoints and analyzed by HPLC for the presence of *p*-cresol and PHBA. In a plasmid-free control *Pseudomonas putida* strain failed to convert 3.3 mM *p*-cresol to PHBA (<0.007 mM). In contrast, *Pseudomonas putida* harboring plasmid pPCU12 produced 0.793 mM PHBA during an overnight incubation (Table 9). PHBA production is, therefore, a new attribute of *Pseudomonas putida* ATCC 29607 when transformed with and expressing *pcu*.

TABLE 9

Strain	Time (h)	PHBA (mM)	*p*-cresol (mM)
control	2	<0.007	3.3
control	5	<0.007	3.4
control	16	<0.007	3.2
pPCU12	2	0.141	2.8
pPCU12	5	0.284	2.8
pPCU12	16	0.767	2.2

EXAMPLE 10

Increased Rate of Production of PHBA from *p*-Cresol by *Pseudomonas mendocina* Harboring a *pcu* Expression Plasmid

Plasmid pPCU12 was transferred by conjugation from *Escherichia coli* S17-1 to *Pseudomonas mendocina* KRC16KDpobA51 as described earlier. Transformants, or a non-transformed control strain, were grown in 15 mL M9 minimal medium containing 1% glucose, 5 mM *p*-cresol, 10 mM $MgSO_4$, tet (15 mg/L) in 125 mL flasks at 30 °C and 225 rpm. Samples were removed at intervals of 2, 5 and 16 h and analyzed by HPLC for the presence of *p*-cresol and PHBA. The *Pseudomonas mendocina* KRC16KDpobA51 strain has a functional chromosomal *pcu* operon, but also has inactivated *pobA* genes to enable PHBA to accumulate. In the presence of the pPCU12 expression plasmid in *Pseudomonas mendocina* KRC16KDpobA51, PHBA accumulates more rapidly to give a concentration of 1.57 mM during the first 5 h incubation, compared to 0.526 mM for the control *Pseudomonas mendocina* strain alone (Table 10).

TABLE 10

Strain	Time (h)	PHBA (mM)	*p*-cresol (mM)
control	2	0.185	3.2
control	5	0.526	2.8
control	16	4.02	0.48
pPCU12	2	0.7	2.8
pPCU12	5	1.57	2.2
pPCU12	16	4.84	0.08

EXAMPLE 11

Production of PHBA from *p*-Cresol by *Agrobacterium rhizogenes* ATCC 15834 Transformed with a *pcu* Expression Plasmid

Agrobacterium rhizogenes ATCC 15834 was grown in nutrient broth at 30 °C and cells harvested during logarithmic growth. The cells were made electrocompetent by washing three times in water, centrifuging at 6000 rpm after each wash. Either the plasmid vector pGV1120 (Leemans et al., *Gene* 19:361-364 (1982)) or pMC4 were electroporated into the cells using 1 mm gap cuvettes at 1.44 kv. Cells were spread on LB plates containing 50 mg/L strep and incubated at 30 °C. Transformants harboring the pGV1120 vector, or the *pcu* expression plasmid pMC4, were grown for 24 h in nutrient broth containing 50 mg/L strep, 10 mM $MgSO_4$, and 1 mM fully-deuterated *p*-cresol. PHBA was extracted from boiled cells with ether and concentrated by evaporation. Gas chromatography/mass spectrometry was used to show that the PHBA formed (1.4 μM) contained 4 deuterium atoms. This experiment proves that it was derived from the *p*-cresol present during culture of the cells.

EXAMPLE 12

Production of PHBA from Toluene by *Pseudomonas mendocina* Transformed with Plasmid pMC3 (*pcu*$^+$ *tmo*$^+$)

Pseudomonas mendocina KRC16KDpobA51 was transformed with plasmid pMC3 and selected on LB + tet (12.5 mg/L) plates at 30 °C. The procedure for cell growth and toluene production was similar to that described in Example 9. The test cultures have 1 mM IPTG present at the growth and PHBA production stages in order to induce transcription from the *lac* promoter. No IPTG was added to the control cultures. Samples were tested for PHBA by HPLC at 1, 2, 4 and 6 h intervals. Table 11 shows that PHBA is produced by induced and non-induced cultures, but with IPTG-treated cells production started earlier, and approached levels that were within the maximum expected based on the amount of toluene added in the flasks.

TABLE 11

	PHBA (mM)	
Time (h)	+ IPTG	- IPTG
1	0.147	0.131
2	0.438	0.180
4	4.985	1.230
6	7.442	4.172

EXAMPLE 13

Generation of Stable *ΔtodC* Deficient *Pseudomonas putida* DOT-T1E Strains

Pseudomonas putida DOT-T1E (CECT 5312) grows on toluene via the toluene dioxygenase pathway (Mosqueda et al., *Gene* 232:69-76 (1999)). The use of this strain for PHBA production from toluene requires its inactivation. In order to generate a mutant deficient in toluene metabolism in DOT-T1E strain, a deletion of the *todC1* gene in the *tod* operon was carried out. Figure 5 illustrates the strategy used and the relevant constructions. The entire DOT-T1 *tod* operon (Mosqueda et al., *Gene* 232:69-76 (1999)) is contained in two plasmids: *todXF* genes borne by pT1-4, and *todC1C2BADEGIHST* genes borne by pT1-125. The approximately 4.5 kb EcoRI/XcaI fragment of pT1-125 which extends from *todC1* to *todD* was cloned at the EcoRI/SmaI sites of a pUC18 Not derivative (de Lorenzo and Timmis, *Methods Enzymol.* 235:386-405 (1994)) that lacked the BamHI and HindIII at the multicopy cloning site to give plasmid pMIR17. The 1.8 kb SspI/EcoRI fragment of pT1-4 containing *todXF* was cloned at the EcoRI site of pMIR17 and the plasmid pMIR20 was obtained. (The unique NotI site present in the SspI/EcoRI fragment was removed before cloning). Most of the *3'*-half end of *todF*, the entire *todC1* gene, and the *5'*-end of *todC2* were removed from pMIR20 as a 1.6 BamHI/HindIII fragment. A 2.2 kb fragment containing the Ω/km cassette (Fellay et al., *Gene* 52:147-154 (1987)), encoding resistance to kanamycin, was cloned at the same position which rendered the pMIR22 plasmid. pMIR30 was obtained as the result of the subcloning in pKNG101 of the NotI fragments, which contained the region corresponding to the *ΔtodC1* and the Km resistance of pMIR22. pKNG101 is a suicide vector in *Pseudomonas* which confers conditional lethality in the presence of sucrose (Kaniga et al., *Gene* 109:137-141 (1991)). pMIR30 was used to replace the *todC1* gene in the chromosome of

Pseudomonas putida DOT-T1E with a deleted version by homologous recombination and a toluene minus DOT-T1E derivative was obtained called ΔtodCkm. The absence of *todC1* gene in the chromosome of the toluene minus isolate was confirmed by PCR with specific primers and in Southern blot.

The stability of the mutant unable to use toluene as the sole carbon-source was tested. The results can be summarized as follows: i) after 90 generations of growth on LB medium under non-selective conditions, i.e. in the absence of antibiotic markers, 100% of cells were resistant to kanamycin and unable to grow in toluene; ii) no growth was observed in M9 liquid minimal medium with toluene as the sole carbon-source; i.e. revertants were undetectable after one week in flasks with 10 mL cultures which had been inoculated with 10^7 CFU/mL); iii) the reversion rate determined as the re-acquisition of the ability to grow on toluene was undetectable (lower than 10^{-9} by the plating technique).

EXAMPLE 14

Cloning of the *Pseudomonas putida pobA* Gene

The *pobA* gene encodes the enzyme *para*-hydoxybenzoate hydroxylase and converts PHBA into protocarechuate. Production of PHBA requires that its metabolism through the pobA pathway be impaired. To this end, pobA was first cloned, then it was inactivated it *in vitro* and the mutation transfered to the chromosome of ΔtodKm. To clone *Pseudomonas putida pobA* gene, a *Pseudomonas putida* KT2440 (ATCC 47054) built in the tetracycline (Tc)-resistant pLAFR3 cosmid (Rodríguez-Herva et al., *J. Bacteriol.* 178:1699-1706 (1996)) was used for the complementation of the *Pseudomonas mendocina* KRC16KDpobA51 (ATCC 55885). In this strain both *pobA* genes are inactivated and so it is unable to grow in *p*-hydroxybenzoate (WO 98/56920). Upon triparental mating with *E. coli* HB101 (pLAFR3::genebank), *E. coli* (pRK600) –a helper strain- and *Pseudomonas mendocina* #303, Tc^R *Pseudomonas mendocina* exconjugants able to grow on *p*-hydroxybenzoate as the sole carbon source were selected. The chimeric cosmids of these clones were isolated, their restriction pattern established and analyzed in Southern blot against the *Pseudomonas mendocina pobA1* gene. A 6 kb BamHI/EcoRI hybridization band common to all cosmids was found and cloned in pUC19 to yield pMIR18. Plasmid pMIR18 was used as the target for "artificial" in vitro random transposition, which was carried out with the Primer Island Transposition Kit (PE Applied Biosystems). A battery of plasmids carrying

the AT-2 transposon at different positions was generated and *E. coli* DH5α cells were electroporated with the heterogeneous mix of plasmids. *pobA* and *pobR* genes were identified by sequencing from specific present at the extremes of the transposable element in pMIR27 (Figure 6). The following primers were used:

1383-1399 oligo pobA1 (+) 5 GCTTCCACGGTATCTCG 3 (SEQ ID NO:104);

1359-1343 oligo pobA1 (-) 5 CAGTCAATCCGCTGCAC 3 (SEQ ID NO:105);

1732-1751 oligo pobA2 (+) 5 GCAGTATGGTCACCTGTTCC 3 (SEQ ID NO:106);

1728-1710 oligo pobA2 (-) 5 GGTTCGACCACCAGGCTAC 3 (SEQ ID NO:107);

1162-1180 oligo pobA3 (+) 5 GGATCTCAAAGCCCTGACC 3 (SEQ ID NO:108);

963-983 oligo pobA4 (+) 5 TGCTGCACAAGGCCGGTATCG 3 (SEQ ID NO:109);

1945-1925 oligo pobA4 (-) 5 GGTCATGAACCAGCTGAAGCG 3 (SEQ ID NO:110);

742-760 oligo pobR2 (-) 5 CCTGTCCGTTAATCGAACG 3 (SEQ ID NO:111).

EXAMPLE 15

Generation of *p*-Hydroxybenzoate Minus Derivative of *Pseudomonas putida* T1-E ΔtodCkm

To knock-out the *pobA* gene in the chromosome of the toluene minus *Pseudomonas putida* ΔtodCkm strain, plasmid pMIR31 was generated with a *pobA* inactivated copy (Figure 6). Plasmid pMIR31 bore *Pseudomonas putida* KT2440 *pobA* gene interrupted by the interposon Ω/Sm. This chimeric plasmid is a suicide vector in *Pseudomonas* and was used as a delivery system for gene replacement of the wild type *pobA* allele for an inactivated copy by homologous recombination.

Pseudomonas putida ΔtodCkm cells were electroporated with pMIR31 and after high voltage pulse, cells were incubated in SOC medium for two h at 30 °C, then centrifuged and the pellet incubated overnight on an LB-agar plate. Finally, Sm-resistant transconjugants were selected on LB plates with Km, 25 μg/mL, and Sm, 150 μg/mL. This selection medium permitted the growth of the clones resulting from a single cointegration event of pMIR31 in the host chromosome, as well as an eventually

successful gene replacement after the resolution of the cointegrate. Two hundred Km^r Sm^r colonies were tested for their ability to grow on *p*-hydroxybenzoate as the sole carbon source and for piperacillin resistance (Pip^r) – the marker of the pMIR31 plasmid that allowed one to confirm the cointegration of the host chromosome of the plasmid. Two of the clones were Km^r Sm^r *p*-hydroxybenzoate $^-$ Pip^s glucose$^+$ was used to confirm the successful allelic exchange of the wild-type *pobA* gene for the inactivated copy confirmed by Southern blot. The double mutant was called *Pseudomonas putida* todCKmpobA::Sm.

EXAMPLE 16

Recruitment of *Pseudomonas mendocina* KR1 Toluene Monooxygenase/*p*-Cresol Utilization Pathways in *Pseudomonas putida* ΔtodCKmpobA::Sm (Construction of a mini*Tn5Tctmo/pcu* Transposon and Production of PHBA)

A transposon was constructed based on a mini*Tn5Tc* with *Pseudomonas mendocina tmo/pcu* genes which permitted integrating these catabolic genes in the chromosome of the double mutant *Pseudomonas putida* ΔtodCKmpobA::Sm and so produced *p*-hydroxybenzoate from toluene. The scheme of the construction of the transposon is shown in Figure 7. The 7.5 kb BamHI fragment of pMC4 containing the *tmoXABCDEF* genes was subloned at the same site in the polylinker of pUC19, generating the plamid pMIR32. The 7.6 MluI/NheI fragment of pPCU17 containing the *pcuRCAB* genes was subcloned at the HindII/XbaI sites of pUC18NotI. In the plasmid so generated, pMIR40, the 7.4 kb BamHI fragment of pMIR32 containing the *tmo* operon was cloned at the *Bam*HI site.

Then the 15 kb NotI fragment containing *pcu* and *tmo* genes was cloned at the unique NotI site of pUT/Tc (de Lorenzo and Timmis, *Methods Enzymol.* 235:386-405 (1994)) generating the plasmid pMIR44 (the unique NotI site of pUT/Tc is located within the transposable element mini*Tn5Tc* born by the plasmid pUT which is suicide in *Pseudomonas*). The transposon was delivered in the chromosome of the double mutant *Pseudomonas putida* ΔtodCKmpobA::Sm via a triparental mating with CC118*λpir* (pMIR44) as a donor and HB101 (pRK600) as a helper strain. Exconjugants Km^r Sm^r Tc^r were selected with a rate of $5x10^{-8}$. The presence of the mini*Tn5Tctmo/pcu* transposon was confirmed in the Tc^r exconjugants by PCR-amplification of the *tmoA* gene. This strain

produces more than 2 g/L PHBA when grown with glucose in the presence of toluene.

EXAMPLE 17

Construction of a *pobA* Mutant of *Pseudomonas putida* KT2440 and Recruitment of mini*Tn5Tctmo/pcu*

Plasmid pMIR31 was also used to replace the wild-type *pobA* gene of *Pseudomonas putida* KT2440 with a mutant allele as it was carried out with *Pseudomonas putida* ΔtodCKm. The resolution of the merodiploid colonies was tested for Sm resistance and Pip sensitivity. One out of 100 colonies exhibited this character and was unable to grow on *p*-hydroxybenzoate as the sole carbon source. The allelic exchange was further confirmed on Southern blot.

The catabolic genes *tmo/pcu* were recruited in *Pseudomonas putida pobA* through a triparental mating with CC118*λpir* (pMIR44), HB101 (pRK600) and *Pseudomonas putida pobA*, as it was previously conducted for the recruitment of the mini*Tn5Tctmo/pcu* in *Pseudomonas putida* Δ*todCKmpobA*::Sm. Nevertheless, this strain only produced trace amounts of PHBA. However production of PHBA was achieved when the regulatory *todST* genes of *Pseudomonas putida* DOT-T1E (SEQ ID NO:112; GenBank Accession Number Y18245; Mosqueda et al., *Gene* 232:69-76 (1999)) were introduced in the strain after subcloning in plasmid pBBR1-MCS5. This strain produced PHBA up to 10-15 mM in 250 mL flasks with 3 mL culture containing about 10^8 cells/mL and incubated at 30 °C on an orbital platform operated at 200 strokes per min. This example indicates that the regulatory genes of the *tod* pathway induce the *tmo* pathway.

The heterologous TodST proteins that control the induction of toluene dioxygenase pathway, are able to induce high levels of expression from the *tmo* pathway genes, and are useful tools to mediate expression of the catabolic *tmo* genes and PHBA production in any organism that does not possess these genes. Previously, Lau and co-workers (*Proc. Natl. Acad. Sci. USA* 94:1453-1458 (1997)) have shown that the two regulatory genes from *Pseudomonas putida* F1, *todS* and *todT*, are members of a two-component signal tranduction family of bacteria uses a histidine-asparate phosphorelay circuit to sense environmental changes. The genes in the instant invention are 95-100% homologous to the tod genes in *Pseudomonas putida* F1.

EXAMPLE 18

Identification of *tmoS* and *tmoT* as Two Components of a Signal Transduction System which Regulates Expression of the *tmo* Operon when Present in either Natural or Recombinant Hosts.

Construction of a *P. putida* strain with a deletion in *todS* and *todT*:

For the generation of the *todST* mutant a derivative plasmid of pKNG101 (Kaniga et al., *Gene*, 109(1):137-141 (1991)) containing a partial deletion of *todST* genes with the insertion of the interposon Ωkm was used. Plasmid pT1-125 (Mosqueda et al., *Gene*, 232(1): 69-76 (1999)) is a pUC18 derivative that bears ~16 kb of the chromosome of *P. putida* DOT-T1E emcompassing part of the *tod* operon of this strain, including the *todS* and *todT* genes (Genebank Accession number Y18245). A 4.3-kb XhoI-HindIII fragment containing the *todS* and *todT* genes was cloned at the compatible SalI and HindIII sites of pUC18 to yield plasmid pT1-155. This plasmid contains a single HpaI site within the *todS* gene and a single EcoRV site within the *todT* gene. The 2.7-kb HpaI-EcoRV fragment was removed from the pT1-155 plasmid and the resulting linear plasmid was ligated to a 2.2-kb blunt-ended fragment carrying a kanamycin resistance cassette obtained from plasmid pHP45ΩKm (Fellay et al., *Gene,* 52(2-3): 147-154 (1987)). The plasmid so generated, pT1-155STKm, was digested with BamHI and the 3.8-kb fragment bearing the corresponding deleted *tod* genes was subcloned within pKNG101 to yield plasmid pT1-ST1Km which carried the insertional deletion Δ*todST*::Km.

The pT1-ST1Km plasmid carrying the *todST*::Km mutation was used to replace the wild type *todST* genes for the corresponding mutant allele into the chromosome of *P. putida* DOT-T1E via homologous recombination. Details of the mobilization of this suicide plasmid, selection of merodiploid strains, and selection of mutants upon allelic exchange were as described before (Ramos-González et al., *J. Bacteriol*, 180(13):3421-3431 (1998)). A clone exhibiting the appropriate replacement of *todST* genes, confirmed by Southern blot and DNA hybridization was selected for further analysis. This mutant was unable to use toluene as the sole carbon source as expected.

P. mendocina tmoS and *tmoT* genes cloned by complementation of the *P. putida todST* deletion strain:

Genes homologous to the *todST* signal transduction system of *P. putida* DOT-T1E were detected in the chromosome of *P. mendocina* KR1 as a 4.5-kb HindIII fragment using *todST* as a probe under low strigency

hybridization conditions (50°C without formamide). A gene bank of *P. mendocina* KR1 containing about 13,000 independent clones was generated by randomly inserting chromosomal DNA fragments in the pLAFR3 cosmid vector using *E. coli* HB101 as a host for transfection by using Gigapack III Gold Packaging Extract (Stratagene #200202). Cosmid pLAFR3 is a lacZ derivative of pLAFR1 (Friedman et al., *Gene*, 18(3):289-296 (1982)). Seven hundred ng of highly intact total DNA of *P. mendocina* KR1, prepared as described (Robson et al., *J.Gen. Microbiol.* 130(7): 1603-1612 (1984)), was partially digested with PstI into 20 kb fragments and ligated at the same site present in 300 ng of PstI linearized cosmid pLAFR3 which had been dephosphorylated with SAP. This library was used to complement the *P. putida* DOT-T1E *todST* mutant with the *P. mendocina* KR1 genes homologous to *todST*. The gene bank was transferred by mobilization to the *P. putida* DOT-T1E *todST* mutant, and transconjugants able to grow on toluene and tetracycline as the marker for the cosmid vector were isolated. All the recombinant cosmids showed positive signals in Southern blots and hybridizations against *todS* and *todT* gene probes, and the same hybridization bands were obtained with both probes. A transconjugant was selected and the cosmid, called pMIR51, showed a 5 kb HindIII hybridization band against the probes. The 5 kb HindIII fragment was cloned in the pBBR1MCS-5 (Kovach et al., *Gene,* 166(1): 175-176 (1995)) broad host range vector and the plasmid so generated, pMAX47-2, conferred the ability to grow on toluene as the sole carbon source after transfer to the *P. putida* DOT-T1E *todST* mutant. The insert of pMAX47-2 was sequenced and the sequence deposited in the Database under accession number AY052500. Two open reading frames of 973 and 220 amino acids, respectively were found. Their translated sequences were compared with all the entries in the non redundant database as described in the BLAST program (Altschul et al., *Nucleic Acids Res.*, 25(17): 3389-3402 (1997)). The proteins showed the highest identity with TodS and TodT, a two component signal transduction system in diverse *P. putida* strains (83% and 85% identity, respectively) (i.e. Ac. No. AF180147, Y18245, U72354). Hereafter the genes cloned in pMAX47-2 will be called *tmoS* and *tmoT*. The above results therefore show that the *P. mendocina* KR1 *tmoST* genes are able to replace TodST in the activation of the toluene dioxygenase *tod* operon.

tmoS and *tmoT* mediated toluene-induction of the *tmo* operon in *P. mendocina* and in heterologous *P. putida* strains:

A transcriptional fusion of the *tmoX* promoter was constructed by using the *'lacZ* wide range reporter plasmid pMP220 (Spaink et al., *Plant Mol. Biol.* 9: 27-39 (1987)). A 7.5 kb-BamHI fragment of pMC4 containing the *tmoXABCDEF* genes was subcloned at the same site in the polylinker of pUC19 and the plasmid so generated was called pMIR32. A 1.2 kb BamHI/ApaLI fragment of pMIR32 containing the *tmo* promoter region was subcloned between the BamHI and HindII sites in pUC19 generating plasmid pMIR36. The 0.8 kb KpnI/PstI fragment of pMIR36 containing the *tmoX* promoter was fused to *'lacZ* in the wide-host range plasmid pMP220. The plasmid so generated, pMIR38, bears the transcriptional fusion of the *tmo* promoter to *'lacZ*, and the role of *tmoS* and *tmoT* on *tmo* expression determined by assay of *lacZ*-encoded -galactosidase.

Plasmid pMIR38 was introduced into different *Pseudomonas* strains to examine the expression of the *tmo* promoter in various hosts. The strains used were *P. mendocina* KR1 (tmoST$^+$), *P. putida* DOT-T1E (todST$^+$), *P. putida* KT2440 (tmoST$^-$ and todST$^-$) and the *P. putida* DOT-T1E *todST* (tmoST$^-$, todST$^-$) mutant. β-galactosidase activities were measured in cells growing in LB in the absence and in the presence of toluene in the gas phase. In the absence of toluene, the expression levels from the P_{tmoX} promoter during growth were below 150 units in all cases. In the presence of toluene, in *P. mendocina* KR1, β-galactosidase activities increased steadily up to 800 Miller Units after 12 hours induction, (6 fold-increase with respect to the basal level in the absence of toluene). In KT2440, no activity was detected as expected since the genes homologous to *todST* and *tmoST* were not present in the chromosome of this strain. The same result was obtained with the DOT-T1E *todST* mutant. The level of β-galactosidase activity in DOT-T1E (pMIR38) increased moderately (4 to 5 fold) in response to toluene. Plasmid pMAX47-2 bearing the *tmoST* genes was able to confer the ability to grow on toluene to the *P. putida* DOT-T1E *todST* mutant. The pMAX47-2 plasmid was introduced into KT2440 (pMIR38) in order to confirm that TmoS and TmoT promoted the transcription from the *tmo* promoter, and high levels of β-galactosidase activity were detected (up to more than 7000 Miller Units), which means a 55 fold induction in the presence of toluene.

The assays reported above were also carried out with all the strains but growing on M9 minimal medium supplemented with 25 mM glucose. It

was observed that during growth in the absence of toluene, very low expression from the P*tmo* promoter was recorded. Strain *P. putida* DOT-T1E (pMIR38), however, exhibited more than 2,000 Miller Units in the presence of toluene (almost a 20-fold induction), indicating the positive influence of *tmoS* and *tmoT* on expression from the *tmo* promoter.